Information is in the eye of the beholder: Seeking information on the MMR vaccine through an Internet search engine

Elad Yom-Tov, PhD[1], Luis Fernandez-Luque, PhD Candidate[2, 3]
[1]Microsoft Research, Herzeliya, Israel; [2]UiT - The Arctic University of Norway, Computer Science Department Tromso, Norway, [3] Norut, Tromso, Norway

Abstract

Vaccination campaigns are one of the most important and successful public health programs ever undertaken. People who want to learn about vaccines in order to make an informed decision on whether to vaccinate are faced with a wealth of information on the Internet, both for and against vaccinations. In this paper we develop an automated way to score Internet search queries and web pages as to the likelihood that a person making these queries or reading those pages would decide to vaccinate. We apply this method to data from a major Internet search engine, while people seek information about the Measles, Mumps and Rubella (MMR) vaccine. We show that our method is accurate, and use it to learn about the information acquisition process of people. Our results show that people who are pro-vaccination as well as people who are anti-vaccination seek similar information, but browsing this information has differing effect on their future browsing. These findings demonstrate the need for health authorities to tailor their information according to the current stance of users.

Introduction

Vaccination programs are one of the greatest public health successes in history. Scourges such as smallpox and poliomyelitis have been completely eradicated or are confined to relatively small areas of the globe. In order to achieve the maximal potential of vaccination campaigns, a large percentage of the population needs to be vaccinated, so as to achieve "herd immunity" which significantly reduces the likelihood of epidemic outbreaks. Sadly, in parallel to vaccination programs, there is a worldwide trend towards a hesitancy to vaccinate, catalyzed by the activism of anti-vaccination groups[1].

The presence of an anti-vaccination movement in a country has been found to correlate with lower vaccination rates[2]. Studies have tracked the presence and importance of those movements in the online sphere, especially social media[3]. This is especially pertinent as nowadays the debate about vaccination is prominent on the Internet[4]. Several studies examined online data sets for the tracking and characterizing the online anti-vaccination movement. For example, Salathe et al.[5] studied the sentiment of people on the social network Twitter with regards to influenza vaccination in the USA. The Vaccine Confidence project is tracking online media to index the sureness of people in vaccine across different countries[6]. Together, these studies show that anti-vaccination content is prevalent and may have a detrimental effect on vaccination rates.

People who wish to make an informed decision on whether or not to vaccinate themselves or their children can therefore find information both supportive of vaccination and opposed to it. Understanding the information seeking process of people as they learn about vaccinations is crucial in order to address their concerns and provide them with convincing information on the importance of vaccination. Many factors affect the information seeking process, and include available sources of information, prior beliefs and understanding, and current news. It has been found that the anti-vaccination misconceptions vary across countries and even across different vaccinations[7]. In the USA, a study found that parents refusing to vaccinate their children were doing so based on information obtained by word-of-mouth, for religious or philosophical reasons, because of low perceived risk of diseases and due to anti-government sentiment[8].

McRee et al.[9] found in a study in the USA that nearly 70% of the parents who had questions concerning vaccinating their child against the Human Papilloma Virus (HPV) searched for information on the Internet. Although that study found that trust of healthcare professionals was high, many people supplement information from these professionals with information sought online.

One of the most common activities online is using a search engine to find information[9]. Internet search engine queries have been shown to reflect both the activities of people in the virtual world[10], as well as those in the physical one. For example, Ofran et al.[11] found a high correlation between the number of searches for specific types of cancer and their

prevalence in the population. Similarly, Yom-Tov and Gabrilovich[12] showed a high correlation exists between the number of prescriptions sold for a drug and the number of people who search for it. Because of this, Internet search engines have been used for learning about medicine and health in a variety of areas. These including knowledge acquisition by cancer patients and their family members[11], obesity and its correlated behaviors[13], and the side effects of medical drugs[12].

Internet search data has also been used to learn about public health, sometimes by measuring effects that are difficult to measure in the physical world. As such, prevalence of influenza was estimated by several researchers[14-16]. More recently, the effect of media reporting of celebrities who are suspected to be suffering from eating disorders on the development of anorexia was studied in Yom-Tov and boyd[17].

In this study we focus on the information acquisition process vis-a-vis vaccines, as evident from the searches of people conducted on a major Internet search engine. To the best of our knowledge there are no studies which examined how people utilize Internet search engines to garner information on vaccination.

Previous beliefs predispose people to a pro- or anti-vaccination stance even before they begin their information acquisition process. This predisposition can lead people who seek information on vaccination to selective exposure, that is, the well-documented tendency of people to people seek information which affirms their viewpoint and avoid information which challenges it[18,19]. However, recent work has shown that, at least in the political domain, provided the information is chosen appropriately, it is possible to cause people to overcome selective exposure and read views opposing to theirs[20]. Our goal is to validate whether selective exposure occurs when learning about vaccination, and if so, to be able in future to use similar approaches to those developed for the political domain in order to better inform parents on the importance and necessity of vaccinations.

Our findings show that when people acquire information about the MMR vaccine they do so with preconceived notions that bias their search for information. Selective exposure occurs both in the information that people choose to seek and in the way they process it. This suggests that providers of information on vaccines need to tailor their content according to the predicted stance of the user.

Methods

The methodology in this work comprises of several stages. First, we devised a method for scoring search engine queries to the attitude of the users who made them towards vaccines. We then applied this score to anonymized data from the Bing search engine. These data were aggregated with vaccine uptake information from CDC and analyzed to show both the impact of individual queries and the temporal progression of users as evident from their queries. In addition, we performed a novel validation of our scoring method using an online advertisement campaign.

Search engine query log data

We extracted all queries to the Bing search engine made by users in the USA during 6 months starting March 2013 which included keywords related to the MMR vaccine. The list of keywords included the vaccine name (MMR or MMRV), as well as the trade names of the vaccine: Priorix, Tresivac, and Trimovax. This resulted in 252,526 queries from approximately 115,714 users. Of these, 9,985 users made five or more MMR-related queries. Data on each query comprised of an anonymized user identifier, time, query text, zip code of the user, the pages displayed to the user as a response to the query, and of these, the pages clicked by users. In order to maintain user privacy, data were anonymized before the investigators had access to them. They were then aggregated prior to analysis, and no individual-level user datum was examined. Thus, all the searches of each individual can be identified, but they cannot be attributed to a specific individual.

Our analysis throughout this paper is based on a scoring of queries made by users and the webpages displayed to them in response to these queries. This scoring is intended to reflect the likelihood that a person making these queries or reading these documents will vaccinate their child against MMR. Broadly, we label a person with a low probability to vaccinate as anti-vaccination, and a person with a high probability to vaccinate as pro-vaccination.

Vaccine attitude scores

We score search engine queries as to the likelihood that a person who made them would be receptive towards vaccination. We refer to these scores as the **Vaccine Attitude Score** (VAS), and in this section detail its calculation.

Our scoring is based on CDC data[1] on vaccination rates (percentage of children vaccinated) in US states and specific urban centers. Each query and page (URL) was scored according to the average of the vaccination rate in the areas from which users to whom the pages were displayed came from. Let z_i be the vaccination rate at the zip code from which the *i*-th query was issued. We score each query and each page d_j displayed in response to this query such that:

$$S_{q_i} = \frac{1}{N} \sum_{i=1}^{N} z_i$$

and

$$S_{d_j} = \frac{1}{N} \sum_{i=1}^{N} z_i$$

Clearly, $0 \leq s_{d_j} \leq 100$. Such a scoring method is akin to the methods used in Yom-Tov et al.[20] to elicit the political intent of queries, using the voting patterns in each zip code. We refer to this scoring method as the VAS. A high VAS implies a higher likelihood that a person issuing the query or visiting the web page will vaccinate. We validate VAS and its use for scoring vaccine-related queries and web pages in the Results, through two separate means.

Results

Assessing the accuracy of VAS for queries and pages

A described in the Methods Section, each query and URL were scored according to the average vaccination rates at the area where people who made these queries reside, so as to obtain an estimate of the likelihood that people reading these pages will vaccinate against MMR. This is similar to the method used in Yom-Tov et al.[20] to score political queries. In this section we validate whether this scoring method, which we refer to as VAS, is accurate in the context of vaccine-related queries scored according to the vaccination rates.

Here we employ two methods for validation. First, we use a novel method utilizing online advertising to capture whether queries are posted by people who are pro- or anti-vaccination. Second, we use human assessors to classify web pages according to whether they perceive the information in them as leading to a higher chance of vaccination.

Estimation of queries using online advertisements

We selected the 10 queries with the highest VAS values, computed according to the methods described in the Methods Section, as well as the 10 queries with the lowest VAS values. For each of these, we placed advertisements on the Bing search engine, such that when people queried for one of these 20 queries, one of two advertisements would be shown at the right hand side of the search results page. The advertisements were shown with equal probability, and the Bing ads system showed them when the price we were willing to pay (our bid) for placing these ads was greater than that of any other advertisement placed for these search terms. In some cases additional, competing ads were shown above or below the advertisements we placed.

Both advertisements were similar in their title ("MMR vaccine"), shape, colors, and the link to which they referred. One advertisement, shown in Figure 1, stated in its text "Do you want to learn about the importance of this vaccine?",

[1] http://www.cdc.gov/nchs/nis/data_files_teen.htm

while the other substituted the word "importance" with the word "dangers". In both cases, clicking the advertisement led people to the CDC page on the MMR vaccine.

We hypothesized that the first advertisement would appeal to people who were pro-vaccination, while the second advertisement would appeal more to people who were anti-vaccination. Thus, we expect that ads on the dangers of the vaccine would be clicked more when they appear next to results for queries made by people with an anti-vaccination stance, and that ads on the importance of the vaccine would be clicked more when they appear next to results for queries made by pro-vaccination people.

The advertisements were shown a total of 5476 times over a period of 15 days. Table 1 shows the clickthrough rate (CTR) for each query and advertisement combination. CTR is the percentage of advertisements which were clicked out of all the advertisements shown. As the table shows, advertisements mentioning the importance of the vaccine tended to be clicked 2.55 times more by people who made queries with high VAS, compared to those by people who made queries with low VAS. Similarly, advertisements which mentioned the dangers of the vaccine were 1.19 times more likely to be clicked by people who made queries with low VAS, compared to those by people who made queries with high VAS.

Thus, we conclude that our results support our hypothesis, that pro- (anti-) vaccination people are more likely to click pro- (anti-) vaccination advertisements, and, furthermore, that our method for scoring queries is accurate in that it assigns high VAS values to queries made more by pro-vaccination people, and vice versa.

Estimation of page VAS using human assessment

Three human assessors were tasked with classifying the 20 pages with the highest page VAS values and the 20 pages with the lowest pages VAS values. Each assessor was asked to label each pages according to the question "Does this page lead to a more pro-vaccination stance?". Possible answers were "yes", "no", or "undecided".

The free marginal Kappa[21] for the three annotators was 0.312, indicating a medium-level agreement. Limiting our analysis to 20 pages which had a standard deviation across users smaller than 2%, because of the need to focus on pages which are accessed by a relatively homogenous population, we find that for pages categorized by a majority of labelers as leading to a more pro-vaccination stance the average VAS was only slightly higher at 72.7, compared to 72.1 (not statistically significant). This minor difference can be for one of two main reasons: Either the page VAS given according to vaccination rates is unrepresentative, or else, the page VAS should be evaluated in the context of how it influenced the next page read.

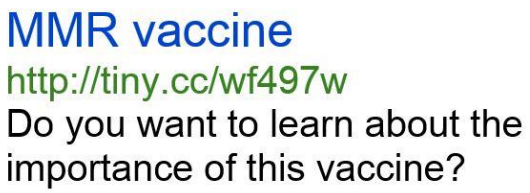

Figure 1. Pro-vaccination ad placed next to Bing searches

	Advertisement	
	Anti	Pro
Low VAS values	0.556	0.468
High VAS values	0.472	1.197

Table 1. Clickthrough rate for pro-vaccination and anti-vaccination advertisements, as a function of query VAS values.

To distinguish between the two, we scored each consecutive pair of pages read by a user, denoted by d_t and d_{t+1} according to the differences in VAS values between them, $s_t = d_{t+1} - d_t$. The average page VAS for those pages labeled as leading to a more pro-vaccination stance was 0.52, compared to a VAS of 0.05 for pages labeled as not leading to a more pro-vaccination stance. Thus, we hypothesize that the latter assumption, that page VAS values should be evaluated according to how they modify a user's reading habits, is the correct one, and that our labeling thus evaluated is a useful way to score pages as to the propensity of users to vaccinate against MMR.

Queries and domains with the highest and lowest VAS

Table 2 shows the list of queries with the highest and lowest VAS in our data. Several interesting observations can be seen from this table. First, both types of queries discuss similar issues, but do so from different viewpoints. For example, the (discredited) link between MMR and autism is addressed as a given in queries with the lowest VAS ("MRR vaccine linked to autism"), but as a possibility in queries with the highest VAS ("link to autism"). Similarly, an Italian court case that suggested that MMR caused autism is searched for in a factual manner in queries with high VAS, whereas people making the queries with a low VAS search for information validating the link ("courts confirm mmr vaccine causes autism"). This suggests that at least some of the people who search for information on vaccines have a preconceived notion of whether or not they intend to vaccinate their children, and are only seeking affirmation of their position. We address this in more detail below.

Among the pages with the lowest VAS, two pages are from CDC, one from Wikipedia, and the remaining are from social media and medical information websites. Pages with the highest VAS include 3 pages from CDC, one from the World Health Organization (WHO), and the remaining are from social media and medical information websites.

In order to obtain a more robust estimation of information sources, we repeated the analysis at the website level, using all websites that had at least 10 scored pages. The score for a website was the average VAS of pages on it. Here, the lowest scored websites included authoritative sources (BMJ, Mayo Clinic, *vaccineinformation.org*, as well as several anti-vaccination websites (*vaccinetruth.org*, *vaccineinjuryhelpcenter.com*), parenting websites (*momtastic.com*), and pharma websites (*merckvaccines.com*). The highest scored websites included parent websites (*mamapedia.com*, *netmums.com*), anti-vaccination websites (*mercola.com*, *vran.org*), government sites (*ny.gov*), and medical websites (*pediatriconcall.com*). Thus, both pro-vaccination and anti-vaccination people read information from similar sources (including government organizations), but, apparently, interpret them differently.

Lowest VAS queries	Highest VAS queries
mmr vaccine linked to autism	mmr vis
side effects of mmr vaccine in adults	when do kids get mmr vaccine
ingredients in mmr vaccine	rash from mmr vaccine
mmr vaccine lot numbers	fever after mmr vaccine
cdc mmr	mmr vaccine administration for adults
mmr vaccine court case	mmr vaccine link to autism
adult mmr vaccine schedule	how to give mmr vaccine
courts confirm mmr vaccine causes autism	mmr vaccine italian court
rash after mmr vaccine photos	mmr vaccine and ppd
mmr vaccination side effects of mmr	vaccine for toddlers

Table 2. Queries with the highest and lowest VAS, according to the vaccination rates in askers geographies.

Modeling the information acquisition

Our goal is to understand what drives a user to read more pro- (or anti-) vaccination information. Therefore, we estimate the contribution of a specific page on the likelihood that a user will read pages with a higher or lower VAS, by modeling the transition of users between pages as a chain of transitions. Let the state of a user at time t_i (i=1,2,...,M) be represented by the VAS of the current page, s_{d_i}, the VAS of vaccine-related pages they read so far, $\sum_{j=1}^{i-1} s_{d_j}$, and the a-priory score of the user, according to the vaccination rate at their locale. The effect of the current page is measured by predicting difference between the VAS of the next page that the user will read and the current page VAS.

The results of a rank regression model, using 14,031 transitions, are shown in Table 3. The table shows that, taken individually, the current page VAS and the average page VAS of past reading have the highest influence on the VAS of the next page. Interestingly, both have a negative effect, that is, the higher the current VAS, the more likely it is to lead to pages with a lower VAS than the current one. This is, possibly, because of interactions between the scores up to time t_i and the score at time t_i. Therefore, model 4 uses all three attributes. Here we find that the R^2 of this model is only slightly greater than that of the model which uses only the current page. However, past reading and user scores contribute to a higher VAS of the next page. The implication of the negative coefficient for page VAS is that many pages are correlated with users reading information associated with a lower vaccination rate. This is an alarming finding, since it suggests that much of the online content on vaccination leads people to read more harmful content.

Since the current page VAS is most strongly correlated with the difference to the next page VAS, we attempt to estimate what in the language of the current page leads to a higher or lower VAS. To do so, we first remove the effect of past reading and user score, by building a rank regression model between these two factors and the difference in next page and current page VAS. Then, we represented each document through its vector space model[22] of words and lexical affinities[23], keeping words that appeared in at least 200 documents, but in no more than a quarter of the documents. We modeled the relationship between the words in the documents and the residual difference using a linear model, constructed using a linear Support Vector Machine (SVM) classifier (with the LibSVM implementation[24], using default settings).

Model number	Variable	Regression coefficient	Model R^2	p-value
1	Page VAS	-0.532	0.283	$<10^{-5}$
2	Past read	-0.387	0.150	$<10^{-5}$
3	User score	0.013	10^{-4}	0.12
4	Page VAS	-0.59	0.297	$<10^{-5}$
	Past read	0.06		$<10^{-5}$
	User score	0.11		$<10^{-5}$

Table 3. Rank regression models of user transitions. Page VAS is the score of the current page read. Past read is the average VAS of all page VASs up to the current time. User score is the vaccination rate at the users' locale.

The 20 words most likely to lead to a higher page VAS are (categorized by the authors, in parenthesis are explanations of the word):

1. Health authorities: guidelines, provider
2. Adverse effect: respiratory, experience (as in "may experience a rash"), avoid (women should avoid becoming pregnant within a month of vaccination), develop
3. Benefits of the vaccination: diphtheria (protection against)
4. Others: answer, lead, review (literature review), personal, scientific, USA, between vaccine, prior, viruses, syndrome, past, started, increased

The 20 words most likely to lead to a lower page VAS are (in descending order of importance, in parenthesis are explanations of the word):

1. Health authorities: control prevention, department, professional, FDA
2. Autism-related: diagnosis (autism diagnosis)
3. Adverse effect: meningitis, swelling, separate (vaccine given in separate parts, rather than all 3 components)
4. Benefits of the vaccination: benefits, pertussis (protection against)
5. Others: issue, activities, problem, means, December, recently, effect, Facebook, put, individual

Thus, both kinds of pages deal with similar topics, but pages leading to lower VAS emphasize adverse effects (and rumored effects) over the more mild symptoms usually experienced.

Differing understanding of similar information

The previous sections suggested that information on the current page and past user reading were most predictive of future reading, and that people who are opposed to vaccination, as well as those who were pro-vaccination read similar information. In this section, we attempt to directly quantify this finding. The null hypothesis is that information on a given page should cause a similar influence on people with differing views, since the information displayed on a page is identical, regardless of a persons' view.

We analyzed the sequence of clicks of each user on pages related to the MMR vaccine, and specifically concentrated on pages that were two or more pages after the beginning of a users' sequence, and at least two pages before it ended. This was done so that the average browsing VAS of a user could be computed both before and after the current click. However, this limited our analysis to 2941 page views, by 979 people who clicked on at least 5 pages.

Users were divided according to whether the average VASs of the pages they clicked on until the current page were above or below the median page VASs. Only 33% of pages were read by people of both groups. This means that people are only 33% likely to read opposing information, and though it is a low percentage, it is higher than that observed for political opinions, which is approximately 20%[20].

Focusing on 108 pages which were read by at least two people from both groups, we measured the fraction of people from each group that read pages with a higher VAS after reading the current page. The Spearman correlation between the fraction of people from each group that read pages with a higher VAS is ρ=-0.004, and not statistically significant. Thus, pages have dissimilar effects on people in the two groups. This means that our null hypothesis is refuted, in that if a page contains information that should lead people to a more positive view of vaccination, the likelihood of reading more positive information differs greatly between people who read pages with lower VASs, compared to those who read higher VASs.

In 78% of pages, the effect was more positive on people who had an average VAS below the median, than on people who had a VAS above it. The association between the page VAS itself and whether it had a bigger effect on one group

over the other is not statistically significant (ranksum test). This implies that, surprisingly, the majority of pages have a more positive effect on people who were previously under the median VAS.

Transitions between VAS quantiles

Our final analysis is concerned with the likelihood of transitioning between page VASs. All page VASs were divided into five quantiles, and the probability of the quantile of the next clicked page given the quantile of the current page was computed. The results are shown in Figure 2. First, as this figure shows, self-loops, indicating that the next page is within the same quantile, are the most common. The average probability of a self-loop is 0.701 (s.d. 0.075). Some of this is due to the coarse division into only five quantiles. However, even when dividing the pages into 20 quantiles, the average probability of a self-loop is 0.607 (s.d. 0.079). Thus, most transitions are to pages with similar VASs.

Figure 2 also shows that transitions are more likely towards quantile 3 than towards quantiles 1 and 5. Indeed, a random walk with restarts[25] (using a random restart probability of λ=0.15) finds that the most likely stable quantile is quantile number 3, with a probability of 0.404, states 2 and 4 have a stationary probability of 0.181 and 0.192, respectively, and stages 1 and 5 have a stationary probability of 0.108 and 0.114, respectively. Therefore, those people who do transition between quantiles are likely to end their search process in pages with VASs in the middle range, indicating that they did not accept extreme views, either towards vaccination or against it.

Discussion

Vaccinations are most effective when the overwhelming majority of the population receives them. Since few countries force citizens to vaccinate, it is important to provide accurate and convincing information to people so as to encourage vaccination. This is a pressing issue especially in the face of anti-vaccination information prevalent on the Internet, which has been linked to a decline in vaccination rates[2].

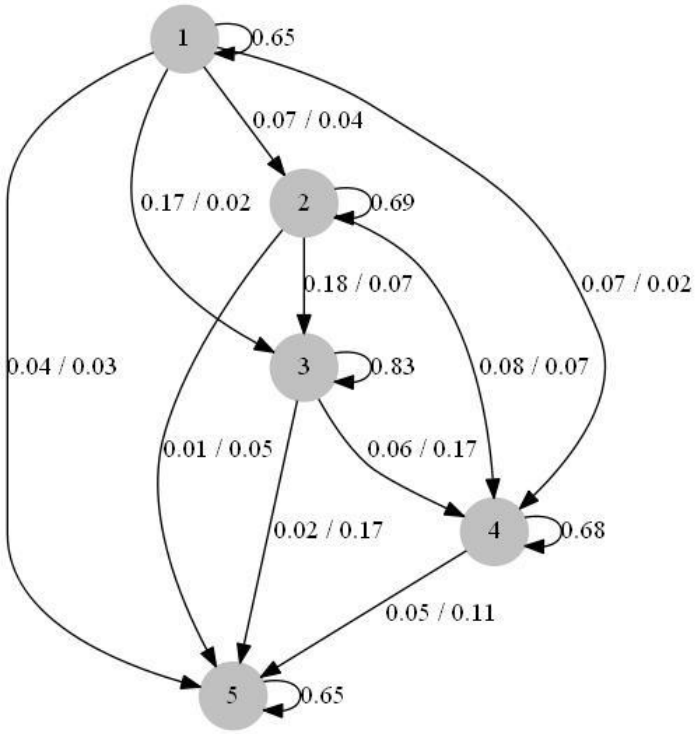

Figure 2. Probability of transition between page VAS quantiles. The left number of each edge represents the probability in the direction of the arrow and the right number the probability in the reverse. Lower quantiles imply lower probability to vaccinate.

In this paper we first developed and validated a way to score queries made to an Internet search engine, and the pages clicked on by users, as to the likelihood that they will be read by people with a pro- or anti-vaccination stance. We used a novel method based on online advertisements to validate our scoring. Our results indicate that our scoring method is useful in identifying such queries, in that people who are pro-vaccination, according to the query VAS value, are significantly more likely to click on advertisements which purport to provide information on the advantages of vaccines, and vice versa. This means that the proposed scoring methods can be used to learn about the information acquisition process, through the observation of sequential page VASs.

When people search for information on the MMR vaccine, their queries reveal their current bias towards vaccines. This represents a problem for pro-vaccination information providers, because their information may not be returned as a result to anti-vaccination queries, and thus pro-vaccination information will not be shown to those people. One possible solution to this problem is to create several types of information, each geared towards people with different attitudes towards vaccination.

Interestingly, though past reading and the information people receive from a web page have statistically significant correlations with future reading, depending on their past reading, people may interpret the same pages differently. Thus, mainstream medical sites feature prominently in the reading of people with an anti-vaccination stance, and the same pages have unequal effects on people of differing stances towards vaccination.

The findings our study have important implications for health authorities, since the Internet is gradually becoming a prominent channel for health education. Firstly, battling the online health misinformation by creating more content is very likely to be suboptimal. The problem it is not the lack of online informative resources about vaccination, but how to design informative resources that are found by those seeking information, whether pro- or anti-vaccination, and providing the most appropriate information for each. Our study provides insights on how to design online information about vaccination. Simple strategies can help, including the usage, within the websites (and their metadata) a vocabulary commonly used by those with anti-vaccination stances. However, further research is required to better understand the reasons for anti-vaccination decisions.

Another approach could be the use of query data for personalizing the health websites. For example, our scores about vaccination stance from search queries can be used to adapt a website about vaccination to provide information which depends on the stance of the visitor. Thus, instead of serving the same content to every person, health providers could offer several versions of their content, each matched to the language and bias of individual users.

Finally, we note that the findings of this study can be further applied to tackle misinformation which has been described by the World Economic Forum as one of the three major threats of the modern hyper-connected society[26].

Acknowledgment

The authors thank Per Egil Kummervold for helpful discussions and assistance in labeling of data.

References

1. Maurice JM, Davey S. State of the world's vaccines and immunization. World Health Organization. 2009.
2. Gangarosa EJ, Galazka A, Wolfe C, Phillips L, Gangarosa R, Miller E, Chen R. Impact of anti-vaccine movements on pertussis control: the untold story. Lancet, 1998;351(9099):356–361.
3. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on youtube. Health Commun, 2012;27(5):478–485.
4. Betsch C, Sachse K. Dr. Jekyll or Mr. Hyde? (how) the internet influences vaccination decisions: Recent evidence and tentative guidelines for online vaccine communication. Vaccine, 2012;30(25):3723–3726.
5. Salathe M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol, 2011;7(10):e1002199.
6. McRee AL, Reiter PL, Brewer NT. Parents internet use for information about HPV vaccine," Vaccine, 2012;30(25):3757–3762.
7. Larson HJ, Smith D, Paterson P, Cumming M, Eckersberger E, Freifeld CC, Ghinai I, Jarrett C, Paushter L, Brownstein JS, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet infect dis, 2013.
8. Fredrickson D, Davis T, Arnould C, Kennen E, Hurniston S, Cross J, Bocchini Jr J. Childhood immunization refusal: provider and parent perceptions. Fam med, 2004;36(6):431.
9. Purcell K. Search engine use. Pew Internet and American Life Project, Tech. Rep., 2012.
10. Goel S, Broder A, Gabrilovich E, Pang B. Anatomy of the long tail: Ordinary people with extraordinary tastes. ACM WSDM, 2010:201–210.
11. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: An analysis of real world data. PloS One, 2012:7(9):e45921.
12. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: Discovery of adverse drug reactions through large-scale analysis of web search queries. JMIR, 2013;16(5): e124.
13. Kuebler M, Yom-Tov E, Pelleg D, Puhl RM, Muennig P. When overweight is the normal weight: An examination of obesity using a social media internet database. PloS One, 2013;8(9):e73479.
14. Eysenbach G. Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance. AMIA, 2006: 244–248.
15. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L. Detecting influenza epidemics using search engine query data. Nature, 2009; 457(7232): 1012–1014.
16. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using Internet Searches for Influenza Surveillance. Clin Infect Dis, 2008; 47(14438).
17. Yom-Tov E, boyd d. On the link between media coverage of anorexia and pro-anorexic practices on the web. Int J Eat Disorder, 2014; 47(2): 196-202.
18. Frey D. Recent research on selective exposure to information. Adv exp soc psychol, 1986;19:41–80.
19. Mutz DC, Martin PS. Facilitating communication across lines of political difference: The role of mass media. Am Polit Sci Rev, 2001;95(1):97–114.
20. Yom-Tov E, Dumais S, Guo Q. Promoting civil discourse through search engine diversity. Soc Sci Comput Rev, 2013.
21. Randolph JJ, Thanks A, Bednarik R, Myller N. Freemarginal multirater kappa (multirater free): An alternative to Fleiss fixed-marginal multirater Kappa. Joensuu learning and instruction symposium, 2005.
22. van Rijsbergen C. Information Retrieval (2nd ed.). Butterworth, 1979.
23. Carmel D, Farchi E, Petruschka Y, Soffer A. Automatic query refinement using lexical affinities with maximal information gain. ACM SIGIR 2002:283–290.
24. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM TIST, 2011;2(27):1-27.
25. Meyn SSP, Tweedie RL. Markov chains and stochastic stability. Cambridge University Press, 2009.
26. Howell L. Global Risks 2013, Eighth Edition. World Economic Forum, 2013. Accessed July 2014 at: http://reports.weforum.org/global-risks-2013/