Journal of Health Communication, 00: 1–8, 2020

ISSN: 1081-0730 print/1087-0415 online
DOI: https://doi.org/10.1080/10810730.2020.1776423

ARTICLE

Algorithms and Health Misinformation: A Case Study of Vaccine Books on Amazon

JIEUN SHIN[1] and THOMAS VALENTE[2]

[1]*Department of Telecommunication, College of Journalism and Communications, University of Florida, Gainesville, Florida, USA*
[2]*Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*

This study examines how vaccine-related books appear on Amazon, focusing on search and recommendation algorithms. We collected vaccine related books that appeared on the first 10 search result pages by Amazon for seven consecutive days and content coded each book. We also collected Amazon's recommendations for each vaccine book and mapped the network of recommendation among these books. First, we found that the number of vaccine-hesitant books outnumbered vaccine-supportive books two to one. Of these vaccine-hesitant books, 21% were written by physicians and medical experts. Second, although we did not find evidence that their search algorithm systematically favored any particular type of book, the three top ranked books across the seven days were all vaccine-hesitant ones. Lastly, using a network model, we found that books sharing similar views of vaccines were recommended together such that when a user views a vaccine-hesitant book, many other vaccine-hesitant books are further recommended for the user. The three most frequently recommended books were vaccine-hesitant ones. The potential consequences of blindly applying commercial algorithms to a complicated health messages such as vaccines are discussed.

The World Health Organization (WHO) has declared *vaccine hesitancy* as one of the top 10 major threats facing the world in 2019. Although the overall immunization rates remain high in the developed countries, pockets of resistance against vaccines exist, causing outbreaks in communities. For example, the U.S. has seen the highest number of the measles cases in 2019 since 1992 (CDC, 2019). Similarly, measles outbreaks associated with low vaccine uptake continue to grow across Europe (WHO 2019).

Vaccine researchers are increasingly interested in the current media environment not only as a window into the phenomenon, but also as one of the potential sources of the problem. Digital media is the primary source of health information for many individuals (HINTS, 2018). Unfortunately, digital media is a space where anti-vaccine messages are spread, and anti-vaccine movements are organized (Basch, Zybert, Reeves, & Basch, 2017; Guidry, Carlyle, Messner, & Jin, 2015; Kata, 2012). Research has shown that exposure to negative information about vaccines can directly and indirectly impact the public's attitudes about immunization (Margolis, Brewer, Shah, Calo, & Gilkey, 2019; Mavragani & Ochoa, 2018).

E-health literacy has been suggested as a potential solution to combat health misinformation (Manganello et al., 2017; Mano, 2014). It focuses on developing individual competencies that enable them to critically evaluate the credibility of online information (Norman & Skinner, 2006). However, comprehensive understanding of the fast-changing, complex media environment is a challenge to e-health literacy. In today's digital worlds, most people rely on search engines and recommendation algorithms to find information and make decisions without fully understanding how algorithms operate (Cohen, 2018).

Therefore, the current study investigates how vaccine content is algorithmically presented to those who are interested in this topic. As a case study, we examine vaccine-related books and the logic of algorithms employed by Amazon. As with other large platform companies, Amazon extensively uses search and recommendation algorithms. Amazon is a particularly useful case for two reasons. First, it provides a unique opportunity to examine an information marketplace that reflects consumers' purchasing behaviors rather than mere browsing. Second, despite widely available free content online, people turn to books as a source for gathering in-depth knowledge about a specific topic (Perrin, 2016).

This study is exploratory in nature as we attempt to understand the internal working of a platform company's algorithms which are often described as "black boxes." We content-coded vaccine related books on the first 10 pages of search results in terms of their stance on vaccines. Our analysis focuses on the rank ordering of books in the search result as well as recommendations. Based on the data collected over seven days, we found that 1) there are over twice as many vaccine-hesitant books than vaccine-supportive books, 2) 21% of these vaccine-

Address correspondence to Jieun Shin, Department of Telecommunication, College of Journalism and Communications, University of Florida, 2066D Weimer, 1885 Stadium Rd, Gainesville, FL 32611. E-mail: jieun.shin@ufl.edu

hesitant books were authored by an authoritative figure in the medical field, 3) other vaccine-hesitant books are recommended once a person views a vaccine-hesitant book.

Literature Review

Algorithm as a Gatekeeper

In contrast to the medical consensus on vaccine safety and efficacy, there is a substantial amount of anti-vaccine messaging spread online (Basch et al., 2017; Kata, 2012; Mitra, Counts, & Pennebaker, 2016). Vaccine critical messages can exert more influence on individuals who are undecided and in the "learning mode." The term "vaccine hesitancy" is increasingly used to recognize a wide range of vaccine attitudes. Narrowly defined, vaccine hesitancy refers to vaccine delay or refusal (WHO, 2019). Broadly defined, vaccine hesitancy also includes those who vaccinate themselves or their children, while still having concerns about their decision to vaccinate (Salmon, Dudley, Glanz, & Omer, 2015). This broader type is particularly vulnerable to misinformation and propaganda potentially being swayed to delay or refuse future vaccines.

According to Thorson and Wells (2015), there are five major categories of actors who influence users' information diet today: journalists, strategic communicators (e.g., public relations), individual media users themselves, social contacts (e.g., friends), and algorithmic filters. Of these gatekeeping actors, the role of algorithms is fast growing as more people rely on algorithms to navigate the Internet. Algorithms can facilitate or constrain the flow of information by displaying a small set of available information and further recommending more information that users might be interested in.

Previous studies have shown that algorithmic output can have a considerable impact on users' attitudes and behaviors (Epstein & Robertson, 2015). In particular, prior research found that a ranking position in the search results determines the visibility of information (Pan et al., 2007; Unkel & Haas, 2017) and the credibility of information (Westerwick, 2013). Furthermore, people tend to prefer advice coming from an algorithm to that of a human (Logg, Minson, & Moore, 2019). For example, users are more likely to purchase a song when they think it comes from an algorithm (Adomavicius, Bockstedt, Curley, & Zhang, 2017).

Auditing the Logic of Algorithms

Despite the increasing power over Internet users' everyday lives, we know little about the logic of algorithms – how they filter information and create recommendations. The algorithms developed by digital platforms are typically proprietary and inaccessible to users. The opaque nature of algorithms has led to a new venue of research, called "algorithm audits." It is an attempt to understand inner workings of algorithms from the outside (Mittelstadt, 2016; Sandvig, Hamilton, Karahalios, & Langbort, 2014).

There are two major functions that algorithms can serve: search and recommendation. The search algorithms find relevant items upon a user's request (e.g., keywords) and prioritize them in the order of importance – defined by each platform. The recommendation algorithms display information that users might be interested in to maximize user satisfaction and increase their interest. These algorithms use various factors which include the user's own past behavior, behaviors of other people who are similar to the user, and the characteristics of the items themselves such as popularity (Cappella, Yang, & Lee, 2015; Smith & Linden, 2017).

In general, the dangers of algorithms are two-fold. First, algorithms tend to harness existing data of varying quality to display and recommend content. Thus, if the input data are biased toward misinformation, algorithms tend to make biased suggestions (Kearns & Roth, 2019). Previous studies (Kata, 2012) have shown that a small number of anti-vaccine activists are disproportionally visible online in comparison to vaccine proponents. Due to a lack of human judgment, algorithms can blindly and unintentionally mirror the loud voice of an anti-vaccine group. This is why some scholars (Gillespie, 2018; Napoli, 2019) call for the development of public-interest minded algorithms.

Second, algorithms have been criticized for their potential to create echo chambers in which users are exposed to one side of the debate, issue, or topic on which they already agree with. Algorithms are optimized to predict what a given user would like to see based on information that they themselves and similar others have consumed. Tufekci (2018) contends that this mechanism reinforces a user's bias and pushes it toward a more extreme view. When this logic is applied to vaccine content, it is possible that an algorithm detects an initial interest from those who have questions about vaccines and suggest more content that are critical to vaccination.

As a case study, we investigate how Amazon's search and recommendation algorithms present vaccine related books. First, we begin by exploring the characteristics of vaccine books displayed on the first 10 pages on Amazon.com in terms of the book's stance on vaccine issues and authors' credentials (RQ1). A few previous studies (Basch et al., 2017; Covolo, Ceretti, Passeri, Boletti, & Gelatti, 2017; Guidry et al., 2015) have examined vaccine messages in various formats such as YouTube videos and social media posts. However, to our knowledge, this is the first study that examines how vaccine issues are communicated in the book format. In general, books are considered to be a voice of experts and perceived to have more "intellectual heft" (Herr, 2017) than other forms of writing (e.g., blogs). Books are also unique in that they typically are not free. Thus, the objective of the first research question is to lay the groundwork for future research by reporting descriptive statistics on the types of vaccine books that are available on Amazon.

RQ1: What types of vaccine books are displayed on Amazon by the search term "vaccine"?

Second, we examine whether there is a systematic bias in rank-ordering books of vaccines. Typically, when a user searches a keyword on Amazon, a ranking system retrieves a list of results that are relevant to the user's query from a corpus of data (collection of available items). Then, the system rank-orders the retrieved items using various features.

Commonly used features include an item's popularity, positive reviews, and price (Diakopoulos, 2019). Examining the ranking of vaccine books will allow researchers to understand the extent to which the ranking algorithm gives visibility to vaccine-hesitant books in comparison to vaccine-supportive ones.

RQ2: What is the relationship between a book's stance on vaccination and its ranking?

Lastly, the study investigates how the stance of a book influences book recommendations. There are two major approaches to designing recommendations for a given user: content-based and collaborative filtering (See Cappella et al., 2015 for review). The content-based approach attempts to recommend items that are similar to those items that a user has previously showed interest in the past. It relies on extensive item descriptions (e.g., book genre, topic) and user profiles (e.g., previous purchases). On the other hand, collaborative filtering seeks to make recommendations based on the behavior of other users with similar interests. Amazon's "customers who bought this also bought" recommendation is a well-known example for collaborative filtering (Smith & Linden, 2017). For example, when a user is viewing an item, Amazon suggests other items that are frequently co-purchased by other users. The goal of this algorithm is to stimulate customers to buy more by exposing them to a set of additional items relevant to their searches.

Some scholars and pundits (Pariser, 2011; Sunstein, 2018) have argued that when these algorithms are applied to political and cultural domains, they can serve as filter bubbles that strengthen exposure to like-minded content and amplify existing biases. For example, examining book-purchasing behaviors on Amazon, one study (Shi, Shi, Dokshin, Evans, & Macy, 2017) observed different genre preferences in the topic of science between two partisan groups. The authors argue that Amazon's recommendation algorithm is in part responsible for reinforcing such partisan dynamics.

Based on the logic of the "customer who bought also bought" algorithm and the findings from previous studies, we can consider two different scenarios. If there is a distinct customer group who co-purchased similar types of vaccine books, then these books (i.e., those sharing the same view of vaccination) would be recommended with one another. Alternatively, if Amazon customers tend to buy vaccine books regardless of a particular point of view, then it is more likely that vaccine book recommendations are made regardless of book stance.

RQ3: Are books that share a similar stance toward vaccines recommended together?

Methods

Data Collection

We first identified books available for sale on the Amazon's book section by submitting a keyword, "vaccine," into the search box and collecting information about books that appeared on the first 10 pages. There were approximately 5,000 books available over 250 pages upon the search request of "vaccine." However, we limit our data collection to the first 10 pages since users rarely go beyond the first few pages (Pan et al., 2007).

We used a web-scraping method for data collection which was repeated for seven consecutive days in March 2018. We performed the queries in the West Coast of the United States. We chose "vaccine" as the keyword due to its dominant use in searching this topic. According to Google Trends data, people search "vaccine" approximately 10 times more frequently than "vaccination" as of October 29, 2018. In addition, an alternative term "vaccination" showed only minor differences from "vaccine" in the search results on Amazon. Although we did not observe the personalization effect of search results, we performed the queries without logging in to the Amazon website and used a browser with no history to prevent any unforeseen potential biases.

It is important to note that we used Amazon's default search setting, that sorts items by the "featured" algorithm. For each book, we collected the book's title, rank in ascending order, unique ID assigned by Amazon, author information, the number of comments, and the average rating for each book. We repeated this procedure for seven days, since it is not known whether search results change considerably on a daily basis.

The number of retrieved books per day ranged between 114 ~ 116 (comprising 164 unique books over seven days). After data collection, we removed books (e.g., fictional books about vaccination) that were not relevant to this study following the procedures explained in detail below. This pruning process resulted in a total number of 81 ~ 86 relevant books each day over seven days (72.1% relevancy on average).[1] This resulted in 104 unique relevant books during the data collection period. We call this dataset "the master list."

Additionally, we collected recommended items displayed under the "customers who bought this item also bought" prompt for each book. We formatted the data as a directional edgelist that contains two columns labeled source and target.[2] A source contains a unique ID of the book of interest, and a target contains a unique ID of a recommended book. The presence of the link indicates the existence of a recommendation from the source book to the target book Since a network requires a boundary, we retained only the items that were present in our master list, constructig recommendation networks among vaccine books.

Coding

Three undergraduate assistants coded each book for the following variables. The first author trained the coders about coding units and coding rules. Over two practice coding sessions, three

[1]Overall, the relevant books overlapped in the range of 72 ~ 96% between a given pair of days during the data collection period. In particular, 45 books appeared throughout all seven days.

[2]Conceptually recommendations should be mutual for two books according to the "bought together" logic. Yet, the algorithm selectively displays "bought together" recommendations for popular items which can result in an asymmetric recommendation network.

coders independently coded a random sample of books (n = 20 each). Subsequently the reliability of their coding was checked.[3] In this process, the codebook was revised based on the feedback from the trainees (e.g., decision to code medical textbooks as irrelevant books).

The coders were provided with a hyperlink to a page on Amazon.com which contains information about each book. Coders were asked to read a description of the book displayed on that page, but not to read the reviews posted for the book. While some pages had detailed descriptions about the book, other pages had little information about the book. If any of the coders indicated that there was not enough information to code the book, we obtained a physical copy for additional information. These cases amounted to 29 books. When we had to purchase a book, we used another online bookstore to prevent any potential influence of our purchases on Amazon's search and recommendation algorithm.

The coders content coded three key variables – relevancy, author's credentials, and book's stance on vaccination. First, they were asked to determine relevancy of each item regardless of whether they were exclusively on child vaccines or vaccines in general. Specifically, the coders were asked to classify a book as "relevant," if the book was concerned about efficacy (i.e., the effectiveness of vaccines in preventing their corresponding diseases) or safety of vaccines (i.e., evidence of harm or risks of vaccination). Otherwise, a book was coded as "irrelevant." For example, irrelevant books included novels involving fictitious characters, books concerned about vaccines for pets, medical textbooks, and immunization log books.

Second, when a book was confirmed as "relevant," we subsequently coded authors' credentials. If the author list included at least one person who was a physician (i.e., Medical Doctor or Doctor of Osteopathic Medicine), a PhD in health or medicine fields, or an official public health organization (i.e., CDC), the author credential was coded as "1," if not "0." According to this rule, all the other author categories such as a PhD in History, JD (Juris Doctor), MBA (Master of Business Administration), and MPH (Master of Public Health) were coded as "0."

Lastly, the book's stance on vaccine efficacy and safety was coded into three categories: "vaccine-supportive," "vaccine-hesitant," and "unclear." In developing our codebook, we followed the guidelines of WHO that defines "vaccine hesitancy" as both "delay" and "refusal" of vaccines. According to this conceptualization, we coded a book as "vaccine supportive, if the book was in favor of vaccines by mentioning vaccine efficacy or safety. In contrast, if a book portrayed vaccines negatively in terms of its effectiveness or safety, the book was coded as "vaccine hesitant." This category included books that were outright against vaccinations as well as those that encourage delay in vaccination or offer an alternative vaccine schedule that has not been approved by the American Academy of Pediatrics (AAP) and the Center for Disease Control and Prevention (CDC) for children. All other books were coded as "unclear," which included history books that summarized controversies surrounding vaccines.

[3]Intercoder reliability measures (Krippendorff's alpha) ranged between 70 ~ 80% initially.

Intercoder reliability was measured with Krippendorff's alpha. Reliability for three variables were high: 91% for relevancy, 95% for credentials, and 84% for stance on vaccination. For each variable, disagreements were resolved by discussion and when consensus was reached it was coded.

Analysis

We started with descriptive statistics to summarize vaccine books. We then examined the differences in the rank across three types of books, using two different measures. We computed the average rank of a book over seven days. Lower score indicates better ranking, thus more visibility. Additionally, we coded whether a book was displayed on the first page at least once during the data collection period. T-tests and a chi-square tests were conducted to examine relationships between vaccine stance and ranking/first page.

Further, we used exponential random graph modeling (ERGM) to investigate recommendation logics among vaccine related books (Robins, 2011). In this study, the network configuration was tested with the R software statnet package (Handcock, Hunter, Butts, Goodreau, & Morris, 2008). In estimating the ERGM, an edges term was included to represent the baseline probability of a tie as well as a term for homophily in book stance (*nodematch*) to test whether books of the same stance are more likely than chance to be connected in the referral network. We also gradually included other factors as control variables such as the stance of the book (*nodeifactor*, tendency of vaccine-supportive books receiving recommendations as compared to vaccine-hesitancy books), the popularity of the book (measured by the number of reviews) and rank of the book (measured by an average ranking position over seven days). We also included endogenous terms such as *reciprocity (*the tendency for them to link each other for recommendations) and *geometically weighted edgewise shared partners* (tendency for two books to share other book recommendations). Separate ERGM models were created for each day's network. Furthermore, Goodness-of-fit (GOF) plots comparing the observed network with a set of simulated networks (n = 100) were checked.[4]

Results

Characteristics of Vaccine Books on Amazon

To answer RQ1, we conducted descriptive analysis of vaccine books. Of the 104 relevant books,[5] 30 books were coded as vaccine-supportive (28.85%), 65 were vaccine-hesitant (62.50%), and 9 were unclear (8.65%). The number of vaccine-hesitant books was more than double the number of vaccine-supportive books. If these books were broken into daily collection, the ratios of vaccine-hesitant books to vaccine supportive books[6] was higher, 3.1 to 1, with an average of 2.8. This indicates that the

[4]The visualization for our model's goodness of fit can be provided upon request.

[5]Some of these were different editions of the same book.

[6]The ratio of vaccine hesitant books to vaccine supportive books for each day was 3.1 (day1), 2.8 (day2), 3.0 (day3), 2.8 (day4), 2.4 (day5), 2.5 (day6), and 2.9 (day7).

number of vaccine-hesitant books was over three times that of vaccine-supportive books. Table 1 summarizes the characteristics of the books in the master list.

Of 104 books, 44 books (42.3%) were written by an author – at least one author if the book had multiple authors – with a relevant credential in the field. Of vaccine-hesitant books, 21.2% (n = 22) were written by an authoritative figure. Vaccine-hesitant books garnered a higher number of reviews (M = 69.7, SD = 136) than vaccine-supportive books (M = 43.3, SD = 141), although the difference was not statistically significant, t(93) = 0.87, p = .39. The average rating of vaccine-hesitant books (M = 4.19, SD = 0.80) was lower than that of vaccine-supportive books (M = 4.96, SD = 3.21), t (78) = −1.71, p = .09.

Search Rankings of Vaccine Books

T-test and chi-square test were conducted to examine RQ2, the relationships between the two types of books (vaccine-hesitant vs. vaccine-supportive) and ranking of search results. Vaccine-hesitant books were ranked higher (M = 56.6, SD = 35.5) than that of vaccine-supportive books (M = 69.4, SD = 37.7), but the difference was not statistically significant (t = −1.60, df = 93, p = .11). The average ranking of unclear books was 64.7 (SD = 37.2). Furthermore, a chi-square test was conducted to examine whether there was association between vaccine-hesitant books and vaccine-supportive books in terms of the proportion of being ranked on the first page of the search result. Consistent with the t-test, the chi-square test indicated that the search result did not differ by the book's stance on vaccination, X^2 (1, N = 95) = 0.005, p = .94.

Recommendation Networks of Vaccine Books

The visual inspection of book recommendation networks (Figure 1 below) shows similar patterns throughout the data collection period. It also reveals that vaccine hesitant books outnumber other types of books and they are clustered together.

Now we turn to the ERGM results. As shown in Table 2 (for Day 1 network), homophily on vaccine stance was significant in all models. Because Model 3 provides a better fit with the observed data, we further examine the coefficients of Model 3 to assess the book recommendation logic. The positive, significant coefficient of recommendation homophily on book stance (b = 1.22, p < .001) suggests that books sharing a similar attitude about vaccines were more likely to be recommended

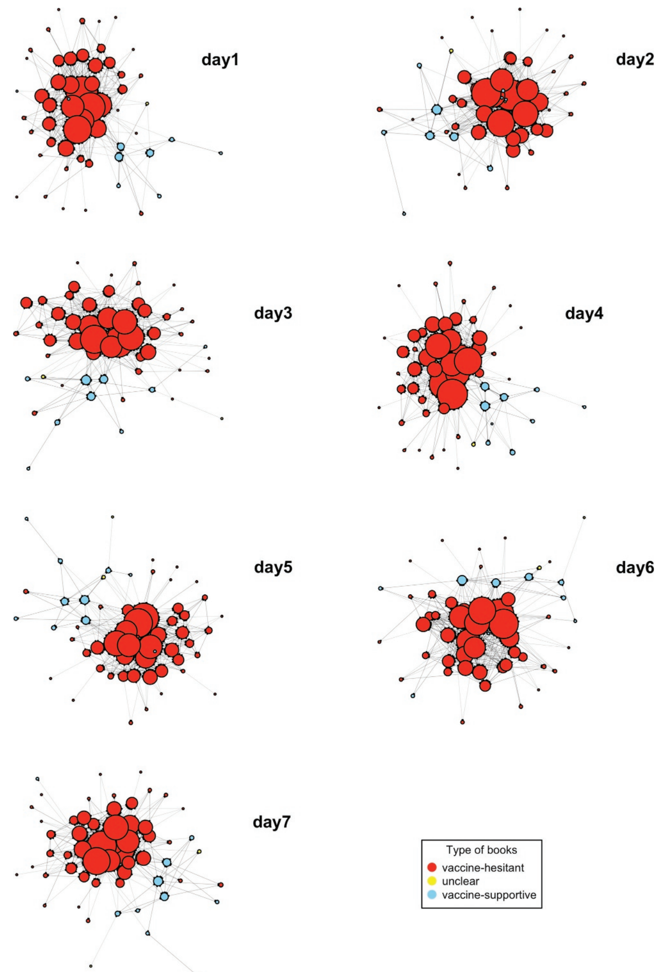

Figure 1. Recommendation networks of vaccine related books. Node size is proportional to indegree (the number of times the book is recommended). Red color denotes vaccine-hesitant books, and blue color indicates vaccine-supportive books. Yellow color indicates unclear category of books.

to one another, after controlling for the book's stance as the main effect, popularity, search result ranking, and other structural tendencies. The odds of similar books being recommended together were over 3 times greater than the odds of dissimilar books being recommended together. Additionally, there was a tendency that books with a higher number of reviews

Table 1. Description of vaccine-related books

	Written by a medical credential (%)	Average number of reviews (count)	Average rating (out of 5 points)
Vaccine-supportive (n = 30)	19.23	43.3	4.96
Vaccine-hesitant (n = 65)	21.15	69.7	4.19
Unclear (n = 9)	1.92	3.11	4.3

Table 2. Day 1 ERGM results (N = 76, MCMC.interval = 5000, MCMC.burn-in = 50000)

Terms	Model 1	Model 2	Model 3
Density	−4.29***	−3.10 ***	−5.74 ***
(*edges*)	(0.18)	(0.21)	(0.36)
Book stance homophily (*nodematch*)	2.55***	2.41 ***	1.22 ***
	(0.18)	(0.20)	(0.15)
Vaccine-supportive books (*nodeifactor*, base = vaccine-hesitant books)		−0.25	0.30*
		(0.21)	(0.14)
Unclear books (*nodeifactor*,base = vaccine-hesitant books)		−1.90	−0.25
		(1.02)	(0.63)
Number of reviews		0.002 ***	0.001 ***
(*nodeicov*)		(0.00)	(0.00)
Ranking position		−0.03 ***	−0.03 ***
(*nodeicov*)		(0.00)	(0.00)
Reciprocity			3.83 ***
(*mutual*)			(0.21)
Triadic closure			1.83 ***
(*gwesp*)			(0.25)
AIC	3087	2557	1764
BIC	3100	2597	1818

*** *p*-value <0.001. SE = standard error. The term for tie density (edges) is used similarly to an intercept term in a regression model. A lower AIC or BIC indicates a better fit.

(b = 0.002, $p < .001$) were more likely to receive recommendation from another book. Furthermore, books ranked higher[7] (b = −0.03, $p < .001$) were more likely to be recommended. Also, book recommendations tend to form a cluster (b = 1.77, $p < .001$) showing a common list of books appearing together. Vaccine-supportive books were slightly more likely to receive recommendations than vaccine-hesitant books (b = 0.30, $p = .04$).

Discussion

Despite the important role of algorithms in our daily life, research on search and recommender systems for health information is scarce. Therefore, we investigated how Amazon algorithmically presents information concerning vaccine safety and efficacy, as an example of platform companies displaying health information. In this study, we focused on ranking and recommendations.

Using content analysis and network analysis, our study found that there were over two times as many vaccine-hesitant books as vaccine-supportive books available on the first 10 search result pages over seven days. Specifically, we found that 63% of books were vaccine-hesitant, whereas 29% were vaccine-supportive. The proportion of these books is an important indicator since they serve as the input data (baseline) for algorithms to produce outputs. It is particularly alarming to find that 21% of vaccine-hesitant books were written by an authoritative figure including MDs and DOs. Previous research has shown that a high degree of medical consensus on vaccines is critical to reducing vaccine hesitancy (van der Linden et al., 2015). The fact that the landscape of this book market is not different from that of social media (Basch et al., 2017; Covolo et al., 2017; Guidry et al., 2015) raises concerns about health science communication.

We did not find a statistical difference in the order of ranking between two different types of books and whether they appear on the first page. This means that Amazon's search algorithm did not favor either type of book, confirming a common belief that the platform algorithm is not designed to maximize public good. Additionally, we found that the three highest ranked books were all vaccine-hesitant. Two of these books were written by pediatricians who offer their own alternative vaccine schedule for children, which is not approved by American Association of Pediatricians (AAP) or CDC. One of them (i.e., Bob Sears, MD) was placed on a 35 month probation in 2018 by the Medical Board of California for inappropriately writing medical exemptions for vaccinations. The other book was written by a self-described "journalist." Yet, the cover of the book was misleading to think that the author was an MD by placing an image of a doctor wearing a white coat. These three books were all highly rated by customers at or above 4.7 out of 5 stars.

Additionally, we analyzed how Amazon recommends potential items of interest with a prompt of "customers who bought this item also bought" using network models. The results show that books sharing the same attitude toward vaccines were recommended significantly more often together than books of dissimilar attitudes. That is, when a user clicks on a vaccine-hesitant book, more vaccine-hesitant books are recommended. This pattern emerged perhaps because initially some people

[7]The small value in the ranking variable indicates the better ranking, thus a negative coefficient.

bought several books that were hesitant about vaccines. Using this data, the algorithm made recommendations which created a feedback loop in which recommendations increased sales of these vaccine-hesitant books. The worst version of this process can be described as an "echo chamber." Digital traces left by some customers reinforces an anti-vaccine theme through a collaborative filtering algorithm.

When we further examined the recommendation networks, the three most frequently recommended books were also all vaccine-hesitant books. Of these three books, two books overlapped with the top ranked books mentioned earlier. The other book was written by an MD who now identifies as a homeopath. The description page states that the purpose of the book is to expose that vaccines are not responsible for the decline in mortality from infectious diseases. The customer review displayed at the top by Amazon's algorithm says that "luckily for our newest son, he has never been vaccinated and truly is healthy" under the review title of "Must read for all doctors and parents."

These findings warrant further discussion on the role of large platform companies in moderating health content. Gillespie (2018) contends that although it is not yet clear what the standards for moderation should be, "we desperately need a thorough public discussion about the social responsibility of platforms" (p.216). Recently, there is a growing demand for developing public-interest minded algorithms to promote societal good (Napoli, 2019). So far, the conversation has mostly focused on political content. This discussion needs to be expanded to the public health arena, given the high-stake issues and the power of platforms. Our findings suggest that the quality of health information displayed by a digital platform is far from what health authorities try to reach.

In addition, e-health literacy educators and practitioners may consider dedicating more time and resources to educating the logic and power of algorithms in today's media environment (Cohen, 2018; Iammarino & O'Rourke, 2018). Previous studies suggested that people tend to follow recommendations given by an algorithm (Adomavicius et al., 2017) and appreciate algorithmic judgment more than human judgment (Logg et al., 2019). Therefore, it is important to provide a broader understanding of algorithms, such as how they filter information, how consumers' data are used for designing algorithms, and what are the unintended consequences. Alongside, media scholars should further investigate complex algorithmic decision-making to increase transparency and update the knowledge required for media literacy.

This study has the following limitations. First, it only examined the vaccine-related books that appeared on the first 10 pages. Content-coding the entire inventory of vaccine books would help researchers to identify a more accurate logic of algorithms. Also, the current study classified a wide spectrum of vaccine-hesitant books into one category. Vaccine-hesitant books could further be classified into books that are outright anti-vaccine and books that encourage delaying and/or skipping vaccines. Additionally, this study is unable to comment on "sponsored" or "best seller" books that usually appear at the top of the research results. Since these labels can greatly influence consumers' interest in the product, future research may want to investigate how vaccine-hesitant books are advertised in the platforms. Lastly, our study is limited in explaining how other factors (e.g., manipulated reviews, sales, and comments) are reflected in the search results, as it focused only on the relationship between the book's stance and the outcome of the algorithm.

Conclusion

Platform companies argue that their algorithms are neutral with respect to content that they are displaying and recommending. Our findings suggest that pre-programmed algorithms may unintentionally channel users' exposure into opinions that are not supported by the science and medical community. This may create the illusion for users that a misinformed minority view is accepted widely in the public. Recently, large platform companies are starting to take some responsibilities by creating an independent oversight body (Facebook) and labeling misinformation (Twitter). Amazon may also need to take a more proactive approach to combating misinformation. In this endeavor, it is essential to raise public awareness about algorithmic filtering and to have a thorough discussion about the role of algorithms in disseminating public health issues.

Acknowledgements

The authors wish to thank the editor, and two anonymous reviewers for critically reading the manuscript and suggesting substantial improvements.

Supplementary Material

Supplemental data for this article can be accessed on the publisher's website.

References

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2017). Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*, *29*(1), 84–102. doi:10.1287/isre.2017.0703

Basch, C. H., Zybert, P., Reeves, R., & Basch, C. E. (2017). What do popular YouTubeTM videos say about vaccines? Child: Care. *Health and Development*, *43*(4), 499–503. doi:10.1111/cch.12442

Cappella, J. N., Yang, S., & Lee, S. (2015). Constructing recommendation systems for effective health messages using content, collaborative, and hybrid algorithms. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 290–306. doi:10.1177/0002716215570573

CDC. (2019). Measles cases and outbreaks. Retrieved from https://www.cdc.gov/measles/cases-outbreaks.html

Cohen, J. N. (2018). Exploring echo-systems: How algorithms shape immersive media environments. *Journal of Media Literacy Education*, *10*(2), 139–151. doi:10.23860/JMLE-2018-10-2-8

Covolo, L., Ceretti, E., Passeri, C., Boletti, M., & Gelatti, U. (2017). What arguments on vaccinations run through YouTube videos in Italy? A content analysis. *Human Vaccines & Immunotherapeutics*, *13*(7), 1693–1699. doi:10.1080/21645515.2017.1306159

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Cambridge, MA: Harvard University Press.

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections.

Proceedings of the National Academy of Sciences, *112*(33). doi:10.1073/pnas.1419828112

Gillespie, T. (2018). Platforms are not intermediaries. *Georgetown Law Technology Review*, *2*(2), 198–216.

Guidry, J. P., Carlyle, K., Messner, M., & Jin, Y. (2015). On pins and needles: How vaccines are portrayed on pinterest. *Vaccine*, *33*(39), 5051–5056. doi:10.1016/j.vaccine.2015.08.064

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, *24*(1), 1548. doi:10.18637/jss.v024.i01

Health Information National Trends Survey. (2018). HINTS 5, cycle 2. Retrieved from https://hints.cancer.gov/data/download-data.aspx

Herr, M. (2017). *Writing and publishing your book: A guide for experts in every field*. Santa Barbara, CA: ABC-CLIO.

Iammarino, N. K., & O'Rourke, T. W. (2018). The challenge of alternative facts and the rise of misinformation in the digital age: Responsibilities and opportunities for health promotion and education. *American Journal of Health Education*, *49*(4), 201–205. doi:10.1080/19325037.2018.1465864

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm–An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, *30*(25), 3778–3789. doi:10.1016/j.vaccine.2011.11.112

Kearns, M., & Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. doi:10.1016/j.obhdp.2018.12.005

Manganello, J., Gerstner, G., Pergolino, K., Graham, Y., Falisi, A., & Strogatz, D. (2017). The relationship of health literacy with use of digital technology for health information: Implications for public health practice. *Journal of Public Health Management and Practice*, *23*(4), 380–387. doi:10.1097/PHH.0000000000000366

Mano, R. S. (2014). Social media and online health services: A health empowerment perspective to online health information. *Computers in Human Behavior*, *39*, 404–412. doi:10.1016/j.chb.2014.07.032

Margolis, M. A., Brewer, N. T., Shah, P. D., Calo, W. A., & Gilkey, M. B. (2019). Stories about HPV vaccine in social media, traditional media, and conversations. *Preventive Medicine*, *118*, 251–256. doi:10.1016/j.ypmed.2018.11.005

Mavragani, A., & Ochoa, G. (2018). The internet and the anti-vaccine movement: Tracking the 2017 EU measles outbreak. *Big Data and Cognitive Computing*, *2*(1), 2. doi:10.3390/bdcc2010002

Mitra, T., Counts, S., & Pennebaker, J. (2016). Understanding anti-vaccine attitudes in social media. *Proceedings of the Tenth International AAAI Conference on web and social media (ICWSM, 2016), Palo Alto,* pp. 269–278.

Mittelstadt, B. (2016). Automation, algorithms, and politics| Auditing for transparency in content personalization systems. *International Journal of Communication*, *10,* 1–12.

Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. New York, NY: Columbia University Press.

Norman, C. D., & Skinner, H. A. (2006). eHealth literacy: Essential skills for consumer health in a networked world. *Journal of Medical Internet Research*, *8*(2). doi:10.2196/jmir.8.2.e9

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: User's decisions on rank, position, and relevance. *Journal of Computer Mediated Communication*, *12*(3), 801–823. doi:10.1111/j.1083-6101.2007.00351.x

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. New York, NY: Penguin.

Perrin, A. (2016). Book reading 2016. Retrieved from https://www.pewinternet.org/2016/09/01/book-reading-2016/

Robins, G. (2011). Exponential random graph models for social networks. In J. Scott and P. Carrington (Eds.), *Handbook of social network analysis* (pp.484–500). London: Sage.

Salmon, D. A., Dudley, M. Z., Glanz, J. M., & Omer, S. B. (2015). Vaccine hesitancy: Causes, consequences, and a call to action. *Vaccine*, *33*, D66–D71. doi:10.1016/j.vaccine.2015.09.035

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, *22,* 1–23.

Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A., & Macy, M. W. (2017). Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour*, *1*(4), 1–9. doi:10.1038/s41562-017-0079

Smith, B., & Linden, G. (2017). Two decades of recommender systems at amazon. *Com. Ieee Internet Computing*, *21*(3), 12–18. doi:10.1109/MIC.2017.72

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton, NJ: Princeton University Press.

Thorson, K., & Wells, C. (2015). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309–328. doi:10.1111/comt.12087

Tufekci, Z. (2018). YouTube, the great radicalizer. *The New York Times*. Retrieved from https://coinse.io/assets/files/teaching/2019/cs489/Tufekci.pdf

Unkel, J., & Haas, A. (2017). The effects of credibility cues on the selection of search engine results. *Journal of the Association for Information Science and Technology*, *68*(8), 1850–1862. doi:10.1002/asi.23820

van der Linden, S. L., Clarke, C. E., & Maibach, E. W. (2015). Highlighting consensus among medical scientists increases public support for vaccines: Evidence from a randomized experiment. *BMC public health, 15*(1), 1207.

Westerwick, A. (2013). Effects of sponsorship, web site design, and Google ranking on the credibility of online information. *Journal of Computer-Mediated Communication*, *18*(2), 194–211. doi:10.1111/jcc4.12006

World Health Organization. (2019). Ten threats to global health in 2019. Retrieved from https://www.who.int/emergencies/ten-threats-to-global-health-in-2019