Modeling U.S. Health Agencies' Message Dissemination on Twitter and Users' Exposure to Vaccine-related Misinformation Using System Dynamics

Shangde Gao
Department of Urban and Regional Planning
Florida Institute for Built Environment Resilience
University of Florida
gao.shangde@ufl.edu

Yan Wang
Department of Urban and Regional Planning
Florida Institute for Built Environment Resilience
University of Florida
yanw@ufl.edu

Lisa Sundahl Platt
Department of Interior Design
Florida Institute for Built Environment Resilience
University of Florida
lisaplatt@ufl.edu

ABSTRACT

This research intends to answer: how do (i) generation frequency and (ii) retweeting count of health agencies' messages impact the exposure of the general users to vaccine-related misinformation on Twitter? We creatively employed a Susceptible-Infected-Recovered (SIR) System Dynamics paradigm to model interactions between message dissemination of 168 U.S. health agencies and proportions of users who are at different exposure statuses to misinformation, namely "Susceptible", "Infected", or "Recovered" status. The SIR model was built based on the vaccine-relevant tweets posted over November and December in 2020. Our preliminary outcomes suggest that augmenting the generation frequency of agencies' messages and increasing retweeting count can effectively moderate the exposure risk to vaccine-related misinformation. This model illustrates how health agencies may combat vaccine hesitancy through credible information dissemination on social media. It offers a novel approach for crisis informatics studies to model different information categories and the impacted population in the complex digital world.

Keywords

COVID-19, misinformation, social media, system dynamics, vaccine hesitancy.

INTRODUCTION

Online health misinformation, i.e., false claims due to the lack of scientific evidence (Chou, Oh, and Klein, 2018), has the potential to affect public health-related behaviors by confusing people's perception and belief towards health issues (Chou et al., 2018). Compared to traditional media, social media platforms have a great potential to propagate health misinformation due to their high reachability, immediacy, and efficiency regarding information diffusion (Jamison et al., 2020). Social media platforms are currently being used widely in crisis events for communicating risk information (Dailey et al., 2018). For example, Twitter has also become a major platform for sharing health information (Hughes and Palen, 2009; Vraga and Bode, 2017) and around 59% of Twitter users in the U.S. have shared and consumed public health information (Wilford et al., 2018). Since the worldwide outbreak of COVID-19, 24.8% of randomly extracted tweets about COVID-19 contained misinformation (Kouzy et al., 2020). Exposure to large-volume health misinformation related to medical therapies can affect beliefs and behaviors of reactive response, which may lead to the rejection of viable disease treatment (Ghenai and Mejova, 2018; Breakwell, 2020). The vaccine-related misinformation has made many U.S. residents trust the link between vaccines and autism (Bode and Vraga, 2018). Recently, a survey conducted by American Nurses Foundation (2020) indicated that 37% of the participants were not confident about the safety and effectiveness of COVID-19 vaccines.

The false beliefs on the safety and necessity of immunization may contribute to the long-existing vaccine hesitancy in the U.S. (Gunaratne et al., 2019). Vaccine hesitancy describes the belief of people who (i) decide to accept vaccination services but have significant concerns about vaccinating their infants; and (ii) delay or reject vaccination services (Dubé et al., 2013). Reluctance to become vaccinated against an infectious disease can decrease safe levels of immunity coverage and increase the risk of vaccine-preventable disease outbreaks and epidemics (MacDonald, 2015). Misinformation spreading on social media platforms can fuel vaccine hesitancy by affecting people's understanding of inoculation performance efficacy (Puri et al., 2020). This type of inaccurate messaging can make the public distrust the guidance from scientific data and medical professionals and make them hesitant towards vaccination services (Yaqub et al., 2014). The public distrust caused by misguided content related to disease inoculations has been observed in the phenomenon of parents being reluctant to immunize their children to vaccine-preventable illnesses (Napolitano et al., 2018).

The increasing infection rate and global spread of SARS-CoV-2 have made it critical to develop a vaccine quickly and to promote vaccination services that ensure robust public acceptance and uptake of immunizations (Coustasse et al., 2021). A notable obstacle affects the widespread vaccination services that the existing belief about the vaccine's negative influence may affect the public's acceptance of the COVID-19 vaccine (National Academies of Sciences, Engineering, and Medicine, 2020). The misguided beliefs in the COVID-19 vaccine efficacy can be related to the vaccine-related misinformation on social media platforms such as Twitter. Based on pre-vaccination-related survey outcomes among U.S. and U.K. residents in September 2020, increasing misinformation about the COVID-19 vaccine appeared to contribute to a decreased intention in becoming immunized when the vaccine was made available among those who would otherwise "definitely" vaccinate by 6.4% in the U.K. and 2.4% in the U.S. (Loomba et al., 2020).

To improve the potential for public acceptance of COVID-19 vaccination, various strategies of combatting misinformation of vaccine effectiveness have been proposed and implemented on Twitter, including validating online information (Al-Rakhami and Al-Amri, 2020) and publishing credible information widely (Vraga et al., 2020). A recent study (Wang et al., 2021b) has investigated the importance of agencies' crises and risk communication during COVID-19. A previous work (Bode and Vraga,2018) also suggested the essential role of the health agencies in providing credible information and moderating the influence of misinformation. A preliminary study (Wang et al. 2021a) has also found that the predominant factual information can suppress misinformation of preventative measures (e.g. wearing masks) in an aggregated manner on Twitter. The ongoing emergence of SARS-CoV-2 infectivity makes it critical to identify proactive strategies that can efficiently and effectively promote vaccine-related credible information.

However, although measures of combatting false contents related to the COVID-19 vaccine have been active, there still appears to be a large group of the population who appear to be communicating distrust for coronavirus vaccination effectiveness (Gallup, 2020; Jamison et al., 2020). Additionally, to date, there is no clear and consistent evidence-based guidance for public health agencies and vaccination services to disseminate messages related to vaccine efficacy that ensures widespread visibility and intent clarity (Dunn et al., 2015; Dailey et al., 2018). To achieve wide public acceptance, it is also important to understand the dynamic relationships between (i) factors of health agencies' message dissemination and (ii) Twitter users' exposure to misinformation in complex environments where the messages and misinformation are being disseminated (Cho et al., 2008). This research intends to answer two research questions (RQ):

RQ1: How does the agencies' daily average frequency of generating vaccine-related messages effect the general users' exposure level to vaccine-related misinformation on Twitter (i.e., the proportion of users who are under the risk of exposure to misinformation)?

RQ2: How does the daily average retweeting count of agencies' vaccine-related messages effect the general users' exposure level to vaccine-related misinformation on Twitter?

We utilize the Susceptible-Infected-Recovered (SIR) System Dynamics model to simulate the dissemination of different information categories about the COVID-19 vaccine (i.e., credible information and misinformation) generated by general Twitter users and health agencies, and the consequential exposure risks of the general users. We define tweet content as credible if it is evidence-based or trustworthy (Fogg et al., 2001). For example, messages published by the selected health agencies and stakeholders are credible and tweets that share similar contents/sentiments are also regarded as credible information. In the contrast, we regard tweets as misinformation if they contain inaccurate or false information (Reuter and Kaufhold, 2018).

To answer the research questions, we measured how generation frequency and retweeting count of agencies' messages influenced the proportion of Twitter users, who were under different exposure statuses, i.e., including "Susceptible", "Infected", and "Recovered" status. These three statuses indicated possible Twitter users' exposure risk to vaccine-related misinformation. Specifically, the users were "Susceptible (S)" if they were potentially "Susceptible" to the risk of exposure to vaccine-related misinformation. "Infected (I)" users, in the contrast, were regarded as ones under the risk of exposure to vaccine-related misinformation at each time step. The users were regarded as "Recovered (R)" if they were reduced from the risk of exposure to the misinformation. We selected the SIR model instead of other variant models (e.g. Susceptible-Exposed-Infectious-Recovered), as we assumed that online users who were exposed to the vaccine-related misinformation were also influenced, so users were under "Infectious" status if they were "Exposed". A baseline scenario for the System Dynamics model was developed by calculating the real-world values of factors in the model based on the Twitter data about the COVID-19 vaccine generated over November and December 2020. This baseline scenario was then adjusted to reflect the changing values of the (i) generation frequency and (ii) retweeting count of health agencies' messages separately. The simulation results were then compared to analyze the population change of Twitter users under the different status of exposure to the vaccine-related misinformation on Twitter under baseline and adjusted scenarios. The results of these simulations offer critical insights into the relationship between the health agencies' message dissemination behaviors and Twitter users' exposure to health misinformation.

METHODOLOGY

Collecting and Retrieving Tweets Relevant to COVID-19 Vaccine

Our preliminary analyses focused on English tweets relevant to the COVID-19 vaccine over November and December in 2020 (61 days in total) when the vaccine development was nearly finished in the U.S.: Pfizer announced successful COVID-19 vaccine candidates on November 9, 2020 (Pfizer, 2020). The U.S. CDC confirmed the vaccine's effectiveness and published orders for utilizing the Pfizer vaccines (CDC, 2021). Additionally, on December 17, 2020, FDA has published a report about authorize the emergency use of COVID-19 vaccines (FDA, 2020). The official information ignited online discussions about the production and the effectiveness of the vaccine in the U.S. Over the study period, we collected tweets with keywords "coronavirus" and "covid" using an open Twitter Streaming API. Then, we utilized "vaccine", "vaccination", and "immunization" as the keywords to filter the collected tweets and retrieved 5,237,303 tweets that are potentially relevant to the COVID-19 vaccine. Because the streaming API could not retrieve the full texts of tweets (most were truncated), we used Hydrator (Documenting the Now, 2020) to extract the full text of each tweet before further text mining.

To remove tweets that contain the keywords but are not relevant in meaning, we utilized the Support Vector Machine (SVM) classifier to further filter the tweets dataset as SVM-based classifier has high performance in similar text classification tasks (Colas and Brazdil, 2006). The annotation was conducted by one researcher and the outcome was agreed upon by the other two researchers. After being trained with the annotated data, the SVM-based classifier has the performance of 95.51% for accuracy, 99.22% for precision, and 95.52% for recall. Finally, we obtained 4,765,945 tweets that are relevant to the COVID-19 vaccine for general Twitter users.

Classifying COVID-19 Vaccine-Relevant Tweets Containing Different Information Categories

The procedure for tweet collection and classification are shown in Figure 1.

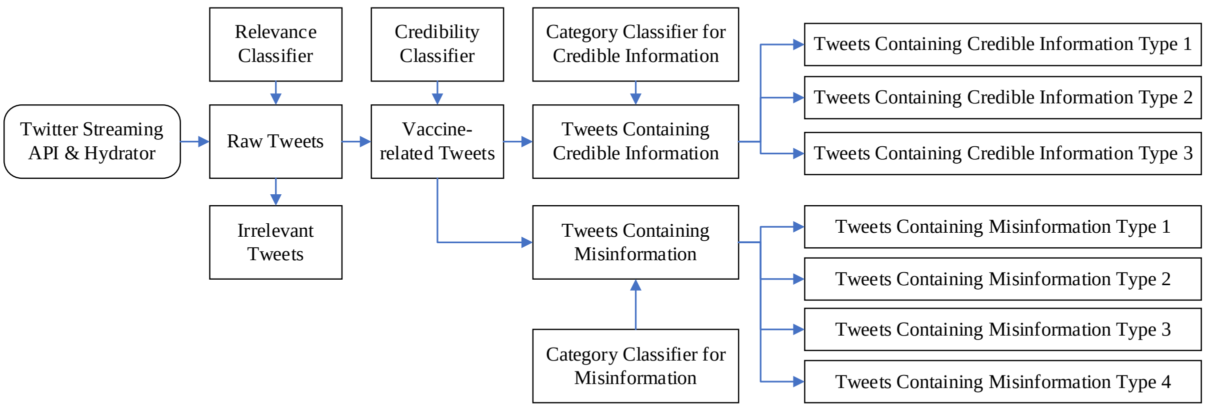

Figure 1. Information Classification for the Collected Tweets

Using the criteria in Table 1, the COVID-19 vaccine-related tweets posted by general users were further classified into two information categories (i.e. misinformation and credible information) with specific types. A tweet was labeled with a specific category/type if one of the criteria was met. One researcher manually labeled 2,000 relevant tweets as the training data and the labeling outcomes were agreed upon by the other two researchers. Using the training data, we first trained an SVM classifier for determining the credibility of tweets, of which the performance is listed in the first row of Table 2. We classified tweets for the overall filtered two-month dataset. We then trained two SVM classifiers to determine the detailed types under credible information and misinformation. After this classification process, we automatically labeled the tweets of normal users based on the criteria they met.

Table 1. Criteria and Examples of Classifying Tweets Containing Different Information Categories

Credibility	Criteria	Example tweets
Misinformation	Type 1: Contain information verified as false by the fact-checking websites (e.g., Poynter)	*"Hydrogels are part of the genome-modifying COVID-19 mRNA vaccines' delivery system and will connect you to the internet."*
	Type 2: Contain the following topics towards the COVID-19 vaccine: negative attitudes towards vaccine creation or influence of COVID-19 vaccine.	*"Packaging of AstraZeneca Covid 19 Vaccine states that the Vaccine contains the cells of aborted male Caucasian babies: Also there is an admission that the vaccine may be harmful or even fatal to humans."*
	Type 3: Contain the contents that dissuade people from getting vaccinated.	*"People may think: oh, there are vaccines coming, right? If COVID-19 is a long-term illness, then NOT getting it is key to protection if vaccines work. If have it? No vaccine can help."*
Credible Information	Type 1: Be verified as true by the verified medical agencies (e.g., CDC, FDA)	*"CDC: a COVID-19 vaccine may not be available for young children until more studies are completed."*
	Type 2: Contain the following topics towards the COVID-19 vaccine: positive attitudes towards vaccine creation or influence of COVID-19 vaccine.	*"CDC Director Robert Redfield tells @DanaPerino: The Covid-19 vaccine is going to begin to be rolled out probably by the end of the second week in December."*
		"Let's dispel some rumors especially because misinformation about COVID-19 may and can cost lives. Enough already! 1. Vaccine efficacy results are real. They were not delayed to hurt or help any politician. 2. The Pfizer vaccine doesnt contain the SARS CoV-2 virus or parts of it!"
	Type 3: Have a basis in scientific fact or valid epidemiological research	*"Pfizer says final data analysis shows Covid vaccine is 95% effective plans to submit to FDA in 'days' https://t.co/nFPqlevDn3"*
	Type 4: Contain the contents that encourage the public to conduct vaccination.	*"Will I get a vaccine for COVID-19 when it is approved and I am eligible to do so? Absolutely."*

Table 2. Performance of SVM Classifiers for Each Step

Classification Step	Accuracy	Precision	Recall
Credibility Classification	91.93%	98.55%	88.31%
Classification of Credible Information Based on Criteria	85.52%	87.17%	83.32%
Classification of Misinformation Based on Criteria	94.48%	92.41%	93.33%

Collecting Risk Communication Messages of Public Health Agencies using a Twitter Search API

To investigate the messages about the COVID-19 vaccine posted by the U.S. health agencies, we have collected an additional dataset to retrieve vaccine-related tweets from the health agencies' official Twitter accounts through the Twitter Search API. We selected 168 vaccine-related agencies as the message sources, including (i) 54 CDC's partners of vaccination services (e.g., American Academy of Family Physicians), (ii) 50 state-level health agencies (e.g., public health agency in Alabama), (iii) 32 state-level hospital association (e.g., California Hospital Association), and (iv) 32 major health systems in the U.S. (e.g., Community Health Systems). These agencies are selected based on the websites of CDC (2016), AHA (2020), and U.S. health systems (Dyrda, 2020). We mainly focused on risk communication (i.e. messages) of federal- and state-level agencies, because vaccination services are governed by the federal and state health agencies in the U.S. They are responsible for promoting vaccination services to control the spread of the COVID-19. Additionally, these agencies also have a high number of followers. For example, U.S. CDC has approximately 3.63 million followers as of February 2, 2021. After the data collection with the Search API, we utilized "vaccine", "vaccination", and "immunization" as the keywords to extract the vaccine-related messages. Finally, we developed a dataset that contains 2,918 messages generated by 168 public health agencies.

Developing A Susceptible-Infected-Recovered (SIR)-Inspired System Dynamics Model of COVID-19 Vaccine Information Dissemination.

We utilized System Dynamics to analyze how the generation frequency and retweeting count of the agencies' information impact Twitter users' exposure to vaccine-related misinformation. The model was generated on the software Vensim PLE (Ventana, 2021). The SIR stock and flow model is shown in Figure 2, which represents how the user population (i.e., stocks) of different exposure statuses changed and transferred (i.e., flow) with other populations. The model structure includes three primary stock variables (which can be measured per day) that illustrate the primary flow of monthly COVID-19 information dissemination on Twitter. We set the time scale as two months, i.e., November and December of 2020, and time unit as one day. A Euler integration was used for simulation. The stock variables include:

Susceptible Twitter User Population refers to the average number of Twitter users observed per day who tweeted about COVID-19 vaccines or retweeted health agencies' messages of COVID-19 vaccine.

Twitter Users Under Infectivity Risk refers to the average number of Twitter users observed per day tweeting with the content that contradicts credible messages being disseminated or retweeting false information regarding COVID-19 vaccine information. We defined "Infected" as under the risk of exposure to vaccine-related misinformation.

Twitter Users Reduced from Infectivity Risk refers to the average number of Twitter users observed per day tweeting with content that was aligned with credible information being disseminated regarding the COVID-19 vaccine.

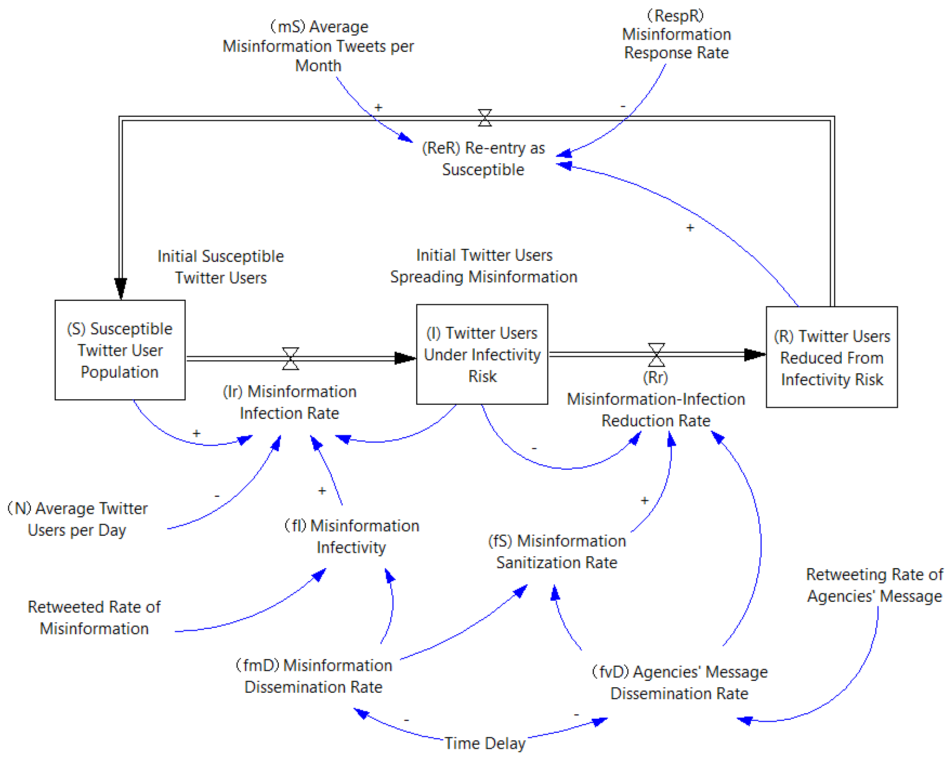

Figure 2. Structure COVID-19 Vaccine Information Dissemination SIR

The development of this model was based on the previous SIR model in the domain of information dissemination (Kabir et al., 2019). We defined users with three statuses: Susceptible (S), Infected (I), and Recovered (R), representing the users' different statuses of exposure to the risk of misinformation of COVID-19 vaccine (see definitions in Introduction). The calculation of transformation rates among these user groups was based on

Equation 1 to 3. “Susceptible” users may become “Infected” with the “misinformation infection rate” (Equation 1) when they were exposed to the vaccine-related misinformation. The infection rate was affected not only by the population of “Susceptible” and “Infected” users but also by the infectivity of vaccine misinformation. The infectivity of the vaccine-related misinformation was influenced by the rate that misinformation tweets were retweeted and the dissemination rate of the vaccine-related misinformation.

$$\text{misinformation infectious risk rate (Ir)} = \frac{\mathrm{S} \times \mathrm{I} \times \mathrm{fI}}{\mathrm{N}} \tag{1}$$

Because of the generation and dissemination of the health agencies’ messages (which were regarded as credible), the Twitter users were “Recovered” from the “Infected” status with the reduction rate (Equation 2). The reduction rate was affected by the population of “Infected” users, dissemination rate of agencies’ messages, and the misinformation “sanitization rate”, which represented the comparison between the number of agencies’ messages and the misinformation spread by persons on Twitter. The sanitization rate was calculated by dividing the dissemination rate of agencies’ messages by that of the misinformation.

$$\text{misinformation infection reduction rate (Rr)} = \mathrm{I} \times \mathrm{fS} \times \mathrm{fvD} \tag{2}$$

However, although the users may be reduced from the risk of exposure to vaccine-related misinformation, it was still possible for them to become “Susceptible” again with the “re-entry rate”, which was calculated by Equation 3. The re-entry rate was influenced by the response rate of vaccine misinformation and the average amount of vaccine-related misinformation tweets in a month, and the population of users who were removed from the “Infected” population also affected this rate.

$$\text{re} - \text{entry rate (ReR)} = \mathrm{mS} \times \mathrm{RespR} \times \mathrm{R} \tag{3}$$

We determined the dissemination rates of misinformation and agencies’ messages based on the collected Twitter data. For the vaccine-related misinformation, the dissemination rate was calculated based on the initial volume of misinformation under specific topics (MisK), the number of misinformation sources (MisN), and the misinformation generation frequency (MisfR). We calculated the MisN by extracting the misinformation tweets generated in the first week and counting the number of all users who posted these tweets. For the agencies’ messages, the dissemination rate was calculated based on the initial average volume of vaccine-related messages generated by each agency (AK), the number of agencies (AN), the messages generation rate (AfR), and the average retweeting count of these messages. The calculation process was shown in Equations 4 and 5. The time delay was also considered in this calculation process, because the misinformation and agency messages generated at a certain time point may affect the population of Twitter users under different risk levels of exposure to misinformation in the future time point. The value of time was set as one week based on the average time gap between the generation time of retweets and original tweets in the collected Twitter dataset.

$$\text{misinformation dissemination} = \mathrm{MisK} \times \mathrm{MisfR} \times \mathrm{MisN} \tag{4}$$

$$\text{agencies}' \text{ message dissemination} = \mathrm{AK} \times \mathrm{AfR} \times \mathrm{AN} \times \text{Retweeting Rate} \tag{5}$$

In the simulation process, we input the initial values for the auxiliary variables of which the values were based on the collected Twitter data, and the definitions of these variables were listed in Table 3. After setting the initial value, we can simulate the population transformation condition under the baseline scenario.

Table 3. Definitions of Input Variables and Variables for Adjustment in the Baseline Scenario

Variable	Definitions/Calculation
Initial “Susceptible” Twitter Users	The “Susceptible” users on the first day.
Initial “Infected” Twitter Users	The “Infected” users on the first day.
Retweeting Count of Misinformation/Credible Information	The proportion of retweeted tweets that contain credible information (or misinformation).
Dissemination Rate of Vaccine-related Misinformation Tweets	Multiplying the initial volume of misinformation of specific topics, the number of misinformation sources, and the misinformation generation frequency.
Dissemination Rate of Health Agencies’ Vaccine-related Messages	Multiplying the initial average volume of vaccine-related messages generated by each agency, the number of agencies, the messages generation rate, and the average retweeting count of these messages.
Generation Frequency of Health Agencies’ Vaccine-related Messages	The average level of daily generation frequency of messages from each health agency.
Retweeting Count of Health Agencies’ Vaccine-related Messages	The average level of retweeting count of all the vaccine-related messages generated by health agencies.

The virtual scenarios allowed us to examine the performance of the implementation of different strategies of disseminating credible information to suppress the vaccination-related misinformation on Twitter. In this study, the adjusted variables (which formed the virtual scenarios) included the generation frequency and retweet count of credible messages generated by agencies, and the observed variables were the population of three groups of Twitter users.

We determined the range of value adjustment for the generation frequency and retweeting count based on the information dissemination behaviors of health agencies in the collected dataset. over November and December 2020, which was also regarded as the baseline scenario. Next, to identify how the value change of generation frequency or retweeting count effect the population of users under different level of exposure risk, we revised the value range for these two variables separately in the adjusted scenarios. After adjusting the values of these two variables in the simulation scenarios, we observed the changing trends and volumes of "Susceptible", "Infected", and "Recovered" populations. The adjustment of the variable's value would be regarded as effective if these changing trends were obvious and heavy.

RESULTS

Baseline Outcome of System Dynamics

We first simulated the information dissemination and Twitter population transformation under the baseline scenario, in which the values of the input variables were listed in Table 4. After the determination of input variables' values, we ran the Vensim PLE to get the population change of "Susceptible", "Infected", and "Recovered" Twitter users, which were shown in Figure 3.

Table 4. Initial Values of Input Variables and Variables for Adjustment in the Baseline Scenario

Variable	Value
Initial "Susceptible" Population	9,810
Initial "Infected" Population	57
Retweeting Count of Misinformation Information	13
Retweeting Count of Agencies' Message	0.69
Dissemination Rate of Vaccine-related Misinformation Tweets	16.18
Dissemination Rate of Health Agencies' Vaccine-related Messages	3.23
Generation Frequency of Health Agencies' Vaccine-related Messages	1.25
Retweeting Count of Health Agencies' Vaccine-related Messages	0.69

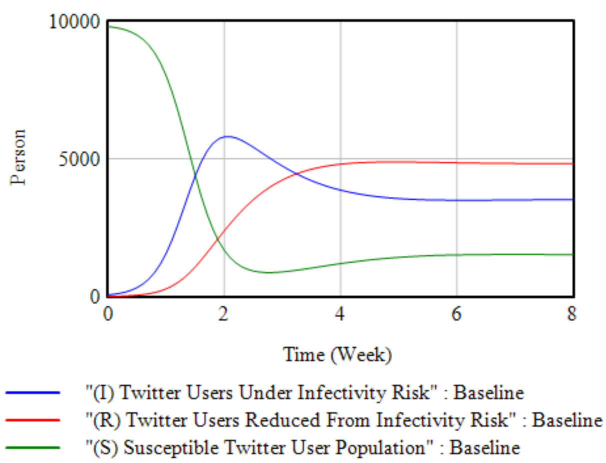

Figure 3. Simulation Outcome of the Baseline Scenario

The baseline scenario showed that because of the dissemination of misinformation and health agencies, the users' population under three exposure statuses showed different changing patterns. In general, the "Susceptible" users affected by vaccine-related misinformation were quickly decreased, and the population of Recovered users was generally increased at the same time. The trend of the population of "Infected" users was different: the population generally increased at the initial stage (i.e., in the first two weeks), then the population slowly decreased to a relatively stable level. The outcomes of the baseline scenario indicated several issues with the current generation frequency and retweeting counts of health agencies' messages. First, at around the second week in the study period, the "Infected" population (higher than 60% of the overall simulation population) reached the peak level of the whole period and was the largest group of simulated Twitter users. If the "Infected" population was predominant, the discussion of the COVID-19 vaccine would be dominated by the users who distrusted the vaccine's effectiveness. Meanwhile, the workload of misinformation correction would suddenly increase and may exceed the agencies' capacity for misinformation correction. Second, although the population of Recovered users became the largest group when the three populations were stable after the third week, the population of "Infected" users was still large, indicating that the health agencies' message dissemination did not effectively help most users avoid the exposure risks. The population change in the baseline scenario provides a reference for the adjusted scenarios, as we can identify whether and how the selected factors (i.e., generation frequency and the retweeting count of health agencies' messages) influence the misinformation prevalence conditions by comparing the population changing trends between adjusted scenarios and baseline scenario.

Simulation of Adjusted Scenarios

In this pilot study, we developed two groups of scenarios by adjusting the generation frequency or retweeting count of the agencies' vaccine-related messages. In each adjustment process, we only change one variable's value and kept the other one stable. The separate adjustments represented that (i) all the agencies more frequently generating vaccine-related messages on Twitter, and (ii) agencies encourage the general Twitter users to retweet their vaccine-related messages. We set the ranges for frequency values and retweeting times (shown in Table 5). The simulation outcomes were shown in Figure 4. The effect of adjusting generation frequency or retweeting count on the changes of three groups of Twitter users was different. After adjusting the two factors separately, we observed the changes of different user populations in the adjusted scenarios focusing on the peak level and the stable population level.

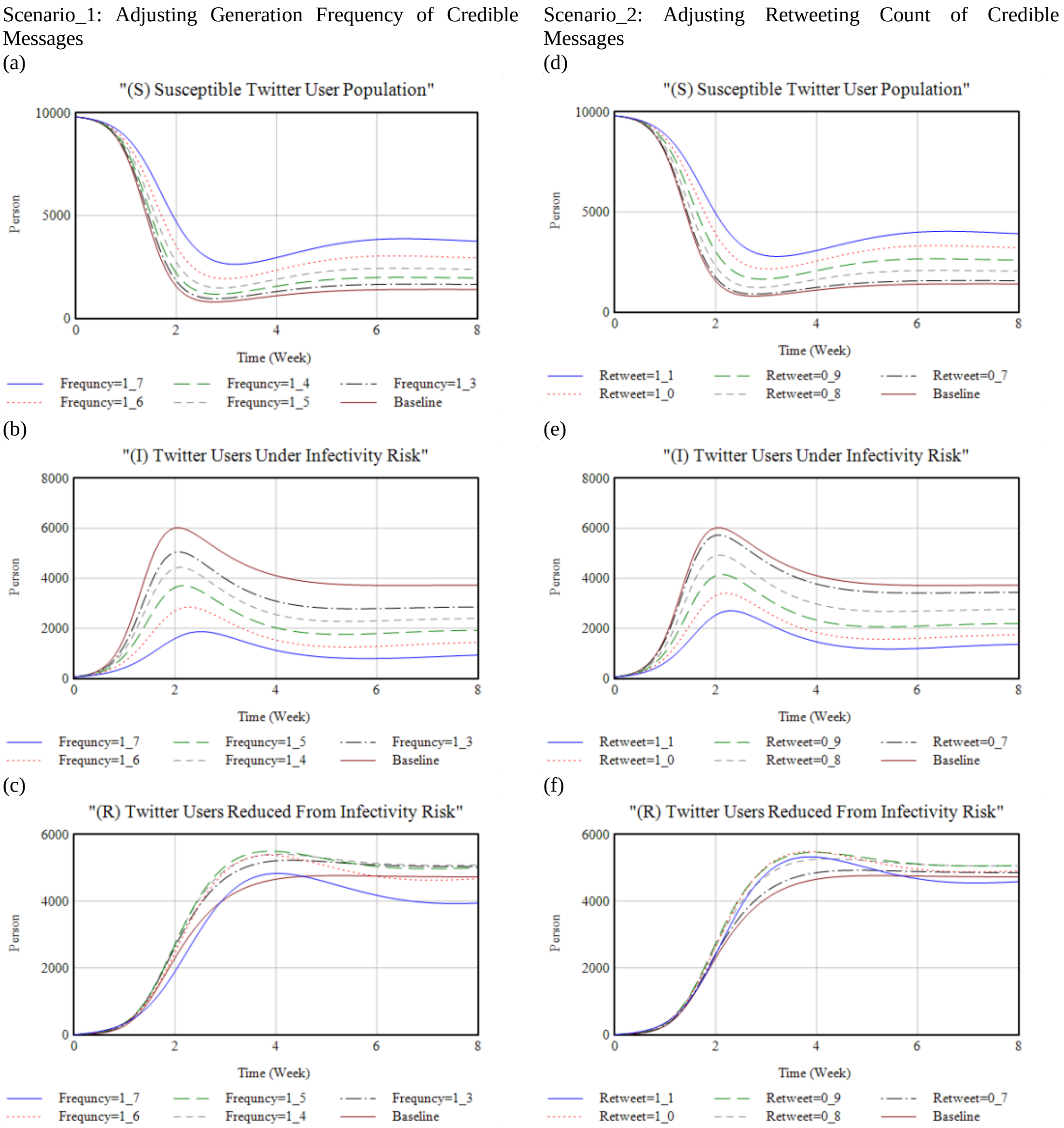

Figure 4. Simulation Outcome of the Adjusted Scenarios

Table 5. Adjustment Range in Virtual Scenarios

Variable	Pilot Values
Retweeting count of health agencies' messages (Times per day)	0.7, 0.8, 0.9, 1.0, 1.1
Generation frequency of health agencies' messages (Tweets per day)	1.3, 1.4, 1.5, 1.6, 1.7

Based on the System Dynamics model in the baseline and adjusted scenarios, we found that augmenting of generation frequency or retweeting count could reduce the possibility of users becoming "Infected" by misinformation through increased exposure to the vaccine-related misinformation. For health agencies, by increasing the message dissemination frequency, the peak and stable value of the "Infected" user population significantly decreased, and more proportion of users were not under the exposure risk of vaccine-related misinformation (RQ1). The adjustment of the retweeting count was still effective, as the "Infected" user population was also suppressed to a lower level comparing to the baseline scenarios. However, comparatively, adjusting the retweeting count was not as influential as increasing dissemination frequency in influencing the "Infected" or "Recovered" user population in the simulation process (RQ2). Also, increasing the generation frequency or retweeting count had little impact on the time when different populations reached peak levels, but the population increased/decreased more rapidly than the change rate in the baseline scenario. This condition indicated that adjusting the generation frequency and retweeting count can suppress the "Infected" or "Susceptible" population to a certain level efficiently. Additionally, we identified that when generation frequency reached a certain value (around 1.5), the effect of adjustment was reduced. Specifically, when the generation frequency exceeded 1.5 (i.e., averagely each agency published 1.5 new messages per day), the peak level and stable level of the "Recovered" population were lower than the condition with lower generation frequency. Such conditions may indicate that there was an optimal value of generation frequency, with which the health agencies can increase the "Recovered" population to the highest available level.

CONCLUSION

This research revealed the impact of agencies' information dissemination on the exposure to health misinformation among the general public. The simulation outcomes of our System Dynamics model provided references for health agencies about social media risk communication strategies and approaches to combat health misinformation. The research advances on previous work in several ways. It innovatively utilizes the SIR-inspired System Dynamics model in a crisis informatics study, which offers a novel approach to model different information categories and their impacts on population exposure to misinformation in online platforms. This model can help identify the impact of credible message dissemination patterns on the overall level of exposure to misinformation on Twitter. The SIR model can inspire more investigations on factors leading to the change among online populations that are exposed to or affected by misinformation. Specific strategies thus can be proposed to mitigate the negative impacts of misinformation by managing these factors. Our research also contributes to bridging several research gaps. Previous studies lacked analysis of vaccine-related agencies' message dissemination behaviors on social media (Chou, Oh, and Klein, 2018). There is also lacking evidence-based guidance for the agencies to disseminate credible messages for mitigating the public's vaccine hesitancy, especially during public health crises (i.e., COVID-19). We focused on the generation frequency and retweeting count of credible messages in this preliminary study. The outcomes can inform agencies and stakeholders of vaccination services to publish credible messages more actively on social media to decrease users' exposure risk to vaccine-related misinformation when vaccination services become more available among the public.

In our future work, we will investigate the effects of content-related factors of the vaccine-related credible messages on suppressing misinformation and consider sentiments in evaluating the impacts of different information categories. Our study period from November and December 2020 in this preliminary work will be extended to cover the periods when the COVID-19 vaccine becomes more available to understand the emergence and evolution of different information categories and effects over a longer period, thereby generalizing the modeling outcomes. We also noticed that the issue of vaccine hesitancy appears to be dissimilar in different countries or communities, and this study only focused on U.S. health agencies and general Twitter users. The communities refer to the population who has similar socio-demographic characteristics. Further work can refine the System Dynamics model to make it area-specific or consider different sociodemographic characteristics, such as determining the range of input variables' values by calculating them based on geo-referenced social media posts generated across countries and communities. We will also improve the System Dynamics by including more detailed components of the information dissemination behaviors on social media to better approximate the complex information and behavior interaction in crises. Additionally, Twitter users cannot represent the whole population due to factors like the digital divide, user preferences, etc. Future studies will investigate strategies to combat health misinformation across different social media platforms over different stages of crises.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2028012. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

AHA. (2020) State Hospital Associations, Retrieved on February 3, 2021, Retrieved from https://www.aha.org/directory/2020-06-09-state-hospital-associations.

Al-Rakhami, M. S., & Al-Amri, A. M. (2020) Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter, *IEEE Access*, 8, 155961-155970.

American Nurses Foundation (2020) Pulse on the nation's nurses covid-19 survey series: covid-19 vaccine. Retrieved on February 3, 2021. Retrieved from https://www.nursingworld.org/practice-policy/work-environment/health-safety/disaster-preparedness/coronavirus/what-you-need-to-know/covid-19-vaccine-survey.

Bode, L., & Vraga, E. K. (2018) See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33, 9, 1131-1140.

Breakwell, G. M. (2020) Mistrust, uncertainty and health risks. *Contemporary Social Science*, 1-13.

CDC. (2016) For Immunization Managers: Partner Websites. Retrieved on February 3, 2021. Retrieved from https://www.cdc.gov/vaccines/imz-managers/partners.html.

CDC. (2021). *Pfizer-BioNTech COVID-19 Vaccine.* Retrieved on April 5, 2020, Retrieved from https://www.cdc.gov/vaccines/covid-19/info-by-product/pfizer/index.html.

Cho, S., Mathiassen, L., & Nilsson, A. (2008) Contextual dynamics during health information systems implementation: an event-based actor-network approach, *European Journal of Information Systems*, 17, 6, 614-630.

Chou, W. Y. S., Oh, A., & Klein, W. M. (2018) Addressing health-related misinformation on social media, *JAMA*, 320, 23, 2417-2418.

Colas, F., & Brazdil, P. (2006) Comparison of SVM and some older classification algorithms in text classification tasks, *IFIP International Conference on Artificial Intelligence in Theory and Practice*, 169-178, Boston, MA.

Coustasse, A., Kimble, C., & Maxik, K. (2021) COVID-19 and Vaccine Hesitancy: A Challenge the United States Must Overcome, *The Journal of Ambulatory Care Management*, 44, 1, 71-75.

Dailey, D., Soden, R., & LaLone, N. (2018) Crisis informatics for everyday analysts: A design fiction approach to social media best practices, *Proceedings of the 2018 ACM Conference on Supporting Groupwork*, 230-243, New York, NY.

Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. A. (2013) Vaccine hesitancy: an overview, *Human Vaccines & Immunotherapeutics*, 9, 8, 1763-1773.

Dunn, A. G., Leask, J., Zhou, X., Mandl, K. D., & Coiera, E. (2015) Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. Journal of Medical Internet Research, 17, 6, e144.

Dyrda, L. (2020) 100 of the largest hospitals and health systems in America | 2019. Retrieved on February 3, 2021. Retrieved from https://www.beckershospitalreview.com/largest-hospitals-and-health-systems-in-america-2019.html.

FDA. (2020)., *Coronavirus (COVID-19) Update: December 17, 2020*. Retrieved on April 5, 2021, Retrieved from https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-december-17-2020.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., ... & Treinen, M. (2001) What makes web sites credible? A report on a large quantitative study, *Proceedings of The SIGCHI Conference on Human Factors in Computing Systems*, 61-68, New York, NY.

Gallup. (2020) More Americans Now Willing to Get COVID-19 Vaccine. Retrieved on December 11, 2020. Retrieved from https://news.gallup.com/poll/325208/americans-willing-covid-vaccine.aspx.

Ghenai, A., & Mejova, Y. (2018) Fake cures: user-centric modeling of health misinformation in social media, *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW, 1-20,

Gunaratne, K., Coomes, E. A., & Haghbayan, H. (2019) Temporal trends in anti-vaccine discourse on twitter, *Vaccine*, 37, 35, 4867-4871.

Hootsuite. (2020). *Digital 2020: Global digital overview.* Retrieved on April 3, 2021. Retrieved from https://wearesocial-net.s3-eu-west-1.amazonaws.com/wp-content/uploads/common/reports/digital-2020/digital-2020-global.pdf.

Hughes, A. L., & Palen, L. (2009) Twitter adoption and use in mass convergence and emergency events, *International Journal of Emergency Management*, 6, 3-4, 248-260.

Jamison, A., Broniatowski, D. A., Smith, M. C., Parikh, K. S., Malik, A., Dredze, M., & Quinn, S. C. (2020) Adapting and Extending a Typology to Identify Vaccine Misinformation on Twitter, *American Journal of Public Health*, 110, S3, S331-S339.

Kabir, K. A., Kuga, K., & Tanimoto, J. (2019). Analysis of SIR epidemic model with information spreading of awareness. *Chaos, Solitons & Fractals*, 119, 118-125.

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E.

W., & Baddour, K. (2020) Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter, *Cureus*, 12, 3, e7255.

Loomba, S., de Figueiredo, A., Piatek, S., de Graaf, K., & Larson, H. J. (2020) Measuring the Impact of Exposure to COVID-19 Vaccine Misinformation on Vaccine Intent in the UK and US, *Nature Human Behaviour*,

MacDonald, N. E. (2015) Vaccine hesitancy: Definition, scope and determinants, *Vaccine*, 33, 34, 4161-4164.

Napolitano, F., D'Alessandro, A., & Angelillo, I. F. (2018) Investigating Italian parents' vaccine hesitancy: A cross-sectional survey, *Human Vaccines & Immunotherapeutic*, 14, 7, 1558-1565.

National Academies of Sciences, Engineering, and Medicine. (2020) Vaccine Access and Hesitancy: Part One of a Workshop Series: Proceedings of a Workshop—in Brief, Washington, DC: The National Academies Press.

Pfizer. (2020a) Pfizer and BioNTech announce vaccine candidate against covid-19 achieved success in first interim analysis from phase 3 study. Retrieved on December 6. https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against

Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020) Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases, *Human Vaccines & Immunotherapeutics*, 16, 11, 2586-2593.

Reuter, C., & Kaufhold, M. A. (2018) Fifteen years of social media in emergencies: a retrospective review and future directions for crisis informatics, *Journal of Contingencies and Crisis Management*, 26, 1, 41-57.

Ventana. (2021) Vensim Software. Retrieved on February 11, 2021. Retrieved from https://vensim.com/vensim-software/.

Vraga, E. K., & Bode, L. (2017) Using expert sources to correct health misinformation in social media, *Science Communication*, 39, 5, 621-645.

Vraga, E. K., Tully, M., & Bode, L. (2020) Empowering users to respond to misinformation about Covid-19, *Media and Communication (Lisboa)*, 8, 2, 475-479.

Wang, Y., Gao, S., & Gao, W. (2021a) Can Predominant Credible Information Suppress Misinformation in Crises? Empirical Studies of Tweets Related to Prevention Measures during COVID-19, *arXiv preprint arXiv:2102.00976*.

Wang, Y., Hao, H., & Platt, L. S. (2021b) Examining risk and crisis communications of government agencies and stakeholders during early stages of COVID-19 on Twitter, *Computers in Human Behavior*, 114, 106568.

Wilford, J., Osann, K., & Wenzel, L. (2018). Social Media Use Among Parents of Young Childhood Cancer Survivors. *Journal of Oncology Navigation & Survivorship*, 9(1), 10-17.

Yaqub, O., Castle-Clarke, S., Sevdalis, N., & Chataway, J. (2014) Attitudes to vaccination: a critical review, *Social Science & Medicine*, 112, 1-11.