An agent-based model about the effects of fake news on a norovirus outbreak

Abstract

Background. - Concern about health misinformation is longstanding, especially on the Internet. *Methods.* - Using agent-based models, we considered the effects of such misinformation on a norovirus outbreak, and some methods for countering the possible impacts of 'good' and 'bad' health advice. The work explicitly models spread of physical disease and information (both online and offline) as two separate but interacting processes. The models have multiple stochastic elements; repeat model runs were made to identify parameter values that most consistently produced the desired target baseline scenario. Next, parameters were found that most consistently led to a scenario when outbreak severity was clearly made worse by circulating poor quality disease prevention advice. Strategies to counter 'fake' health news were tested. *Results.* - Reducing bad advice to 30% of total information or making at least 30% of people fully resistant to believing in and sharing bad health advice were effective thresholds to counteract the negative impacts of bad advice during a norovirus outbreak. *Conclusion.* - How feasible it is to achieve these targets within communication networks (online and offline) should be explored.

Keywords: Agent-based-models; outbreak; norovirus; fake news; filter bubbles.

Un modèle basé sur les agents sur les effets de information fallacieuse sur une épidémie de norovirus

Position du problème. - La désinformation sur la santé est une préoccupation de longue date, en particulier sur Internet. *Méthodes*. - À l'aide de modèles à base d'agents, nous avons examiné les effets de telles informations erronées sur une épidémie de norovirus, ainsi que certaines méthodes permettant de contrer les effets possibles de «bons» et de «mauvais» conseils en matière de santé. Le travail modélise explicitement la propagation de la maladie physique et des informations (en ligne et hors ligne) comme deux processus distincts mais en interaction. Les modèles comportent plusieurs éléments stochastiques; Des répétitions de modèles ont été effectuées pour identifier les valeurs de paramètre qui produisaient le plus systématiquement le scénario de base cible souhaité. Ensuite, il a été trouvé des paramètres qui conduisaient systématiquement à un scénario dans lequel la gravité des épidémies était clairement aggravée par la diffusion de conseils de prévention de maladies de qualité médiocre. Des stratégies pour contrer les «fausses» nouvelles sur la santé ont été testées. *Résultats.*- Réduire les mauvais conseils à 30% du total des informations ou rendre au moins 30% des personnes totalement réticentes à croire en des mauvais conseils sur la santé et à les partager est un seuil efficace pour contrecarrer les effets négatifs d'un mauvais conseil lors d'une éclosion de norovirus. *Conclusion*. - La possibilité d'atteindre ces objectifs dans les réseaux de communication (en ligne et hors ligne) doit être explorée.

Mots Clés : Modèles à base d'agents, épidémie, norovirus, infox, bulles de filtres

1. INTRODUCTION

Political campaigns in 2016 sparked interest in 'fake news', a term with no fixed definition [1]. At its most pernicious, it can mean mostly or entirely false information, often deliberately false or at least created with no regard for truth, yet purporting to be entirely truthful, and therefore indisputably unhelpful when trying to make informed decisions [2, 3]. Worry that fake news might be used to distort political processes or manipulate financial markets is well-established [3-6].

Less studied is the possibility that misinformation spread could harm human health, especially during a disease outbreak. Accurate information spreading during epidemics that generates more protective behaviour, as well as other potential behaviour responses (usually beneficial) following increased awareness of disease prevalence have been widely modelled, reporting typically on how disease dynamics might change as a result (usually resulting in improvements to human health outcomes). But fewer if any studies have tried to model behaviour response that might affect human health during an outbreak that is linked to dangerously *wrong* information [7].

We built models that capture the impacts in response to spread of *dangerously misleading information*, which we simply call *bad advice*. The premise of the modelling is that some types of information about a disease or outbreak ("bad advice"), if truly believed, would lead to people taking fewer or less effective protective measures. Examples of riskier behaviour would be increased physical contact, less hand-washing, less disinfection, or more indirect physical contact such as sharing food or with contaminated fomites. We were interested in gastro-intestinal illnesses, which are rarely considered in individual-based models for infectious disease [8]. Norovirus is the most common GI bug worldwide [9] including in the UK [10]. It can overwhelm health services [11-15]. For modelling purposes, norovirus is convenient because of short duration, familiarity unlikely to cause flight response, and very rare death. This modelling suited the environment of an agent-based model (ABM) that simulated physical contact that could transmit disease alongside information sharing that did not require physical contact.

2. METHODS

2.1 Overview

The model imagined a strain of norovirus for which there was no prior immunity. We incorporated observed parameters where possible, and for UK if required to be very specific. Otherwise, parameters and assumptions were adjusted empirically to yield desirable performance metrics, as described below. The key behaviour response was taking effective precautions (TP). TP does not mean a specific single behaviour (such as reducing contact, not sharing food, washing hands, disinfection, etc.). Rather, TP is meant to be an umbrella term (expressed numerically as a percentage) that includes *all behaviours* that could effectively prevent disease acquisition or transmission. TP describes behaviour when contact could be made with someone with active known disease; we don't consider precautionary behaviour in absence of circulating disease.

The modelling stages are shown in Fig. 1. First, we designed a stage 1 scenario for a disease outbreak, where disease acquisition was partly dependent on individual precautionary behaviour that was static and unchanging in stage 1. A mean TP value was found that reliably yielded our target r0 after many iterations (required due to the random-probabilistic design of models). The next stage (2) model had multiple social network and information sharing attributes parameterised by real world observations and established theories. In stage 2, a 40% increase in the r0 value was achieved (compared to stage 1), creating a scenario where circulating bad information led to greater person to person spread. Stage 3 considers two intervention strategies to counter the impacts of 'fake news' on health protection behaviour. Additional items S1 and S2 provide further details about model construction. At least 100 simulations ran to test parameter values in each stage model. The key outbreak measures reported were: r0, overall attack rate, peak prevalence and outbreak duration.

2.2 Stage 1: SEIR Model without information spread that changes behaviour

We wrote a susceptible-exposed-infected-recovered (SEIR) model in Netlogo [16]. The world shape was a torus (eg. going off the bottom means re-entry at the top), with visible area measuring 88x90 patches that agents can move around on. Initial agent location on the grid was quasi-random. The model has universal 8-hour duration night-time periods when all movement stops and new contacts do not occur. Night-time was explicitly modelled because norovirus has a relatively short incubation period and duration of illness; both about 36 hours [9, 17-19].

Disease incubation periods and recovery-times were assigned individually to each agent from a random-normal distribution. Both attributes had target mean = 36 hours but with additional desired features for the distribution of their values, as shown in Table 1, to conform with data reported in relevant literature. Agents were assumed to only be infectious to others during active illness. The model was initialised with many agent-own attributes (Table 1).

Agents were spatially distributed in small clusters with a like-minded attitude (the reject-est attitudinal trait, as described below). These clusters often spatially overlapped. Empirically, we found that 1600 agents achieved the target mean contact rate expected for the UK (11.74/day [20]) in non-outbreak conditions. Time steps were hours; the model starts at 7am on the first morning. A start-time was important to set sleep periods (when new contacts paused but infection and incubation would continue). Agents return ‘home’ every evening at 11pm. Each well agent moved in a random direction one step in the agent-world each hour; ill agents moved 0.2 steps. The agent world space is not to scale with the real world. Rather, each movement represents time-space; opportunities for potential new contacts due to travel (by any means). 2% of (randomly selected and located) agents were infected at the start of each simulation.

The baseline mode had a mean basic target reproduction number from community outbreaks (r0) =1.9; [21]. In real life, whether disease is contracted can depend on three factors: separate probabilities of either susceptible or infectious person taking adequate precautions, as well as the amount of shed virus. In reality, these components are hard to observe or separate. Model infection risk could be captured in a single global infection-chance parameter, but in our model, risk-taking behaviour of susceptible and infectious persons had to be distinct, so that the likelihood of unsafe behaviour could vary individually and over time. Each agent needed a “take precautions” (TP) property, to represent the probability of taking effective precautions to avoid transmission, given unobserved and not parameterised amount of viral shedding. Thus, TP was individually assigned to agents according to a probabilistic distribution, constrained to range 1-100%, with a pre-specified population mean and assumption of normality around the mean. TP values are highly influential in the model and easily alter the basic reproduction number (r0). Stage 1 is the phase of our modelling where we use multiple iterations to establish the mean population TP value that most reliably led to the target r0 (1.9). The stage 1 model tests candidate mean TP values from 70-90% (in increments of 0.1-1%; standard deviation = mean/4. Potential changes in take-precautions (TP), due to circulating advice, *is the key behaviour response* in our stage 2-3 models, as described in subsequent sections where individual TP values vary in response to circulating advice.

Infection was transmitted when infectious agents encounter susceptible agents and neither took adequate precautions to avoid transmission (tested stochastically and hourly). Incubation and illness durations were determined stochastically (with mean = 36 hours). Recovered individuals were immune. Many features that could more ideally replicate norovirus outbreaks were not included, such as shedding of virus post-illness, increased transmission due to closer night-time contact, environmental and foodborne transmission. These were omitted to reduce model complexity and instead focus on the impacts of information spread.

Any TP value ever set to < 0 was reset to 0 while values > 1 became 1. The model ran until no one was incubating or infectious.

2.3 Stage 2. Incorporating Information spread and how that could change behaviour during an epidemic

Advice is information that may be true or false by objective standards. *Good advice*, if believed, encourages taking precautions that will be effective. *Bad advice* in our models is information that promotes not taking effective precautions or other behaviour that increases risk of transmission. Misinformation online is typically much more exciting than true information [22] False stories are observed to be more surprising and novel than true stories, and more likely to have counter-hegemonic framing [23]. We therefore assume that bad advice elicits stronger emotions, and often challenges orthodox or 'mainstream' sources. These attributes make bad advice attractive and thus often shared with others. Believing bad advice could mean increased physical contact, more intimate types of contact, less hand-washing, less disinfection, sharing food or touching contaminated fomites: effectively, taking fewer precautions to avoid disease.

Our stage 2-3 models assume that taking-precautions (TP) changed in response to each exposure to bad or good advice. No existing data suggested the magnitude of change after each information exposure. It was most useful to find a TP change that consistently led to a *worse* outbreak (we defined "worse" = 40% increase in r0, from 1.9 to 2.66). Therefore, we repeatedly tested many change values to find one that most consistently led to r0 = 2.66 (see "Finishing stage 2" below).

How the take-precautions attribute changed in response to advice depended on trust in information sources.

2.4 (Dis)Trust in The Establishment

Distrust in conventional authorities is closely linked to tendency to believe in conspiracy theories (CTs) [24]. The best predictor of belief in a specific CT or domain of CTs is pre-existing belief in a CT [25, 26]. CTs are relevant to believing bad advice, because fake news stories often use conspiracy theories to allege that conventional advice or conflicting information should be disregarded. CTs are also incorporated into fake news to increase circulation [23, 27, 28].

Predisposition to distrust establishment sources exists within our model as a stochastic property assigned individually to each agent called *reject-est* ("reject establishment"). Reject-est ranges from 0 to 1. Reject-est affects likelihood of sharing information as well as predisposition to change behaviour in response to bad advice (assumed to be both more emotionally framed and contextualised with counter-hegemonic bias, making the information more attractive to those with a high reject-est bias).

Table 1. Key agent-own parameters

Variable or feature	Purpose	Where used in model(s)	Allowed range	Plausible or likely values	Other info or assumptions
members-my-bubble	List of agents that comprise each person's information "filter bubble" [26]	Who advice is shared with chosen from this bubble.	Size = approx. 80-230, to conform with Dunbar# expectations, mean ~150	See Dunbar# research [29]	Item S2 explains in more how bubble membership was constructed.
recovery-time:	Indicate duration of infectiousness and illness (assumed to perfectly coincide)	To decide transition from infectious to recovered status.	1 to 2x mean = 36 hrs; 1 hr min.	24-72h [9]	36h treated as population mean
reject-est: tendency to believe fake news and reject "establishment" (conventional) messages	(%/100) how much they tend to believe bad advice.	Used for likelihood of sharing types of information and predisposition to changing behaviour, always considered in comparison to mean group reject-est value. Relates to var=take-precautions, and how much agents are influenced by types of advice.	0-1; higher means accept BA more easily	38.88% is supported by the literature.	Does not change during outbreak
take-precautions (TP)	Likelihood of taking precautions to prevent getting disease	Set at start from distribution with mean = 79.6%, SD = 19.23%; which consistently yielded target r0 = 1.9. TP indicates the % of contact moments when agents take effective precautions	0-1	Scaled 0-100%	TP changes during outbreak, in response to advice exposed to (stages 2-3)
time-to-incubate	To indicate when agent changes from incubating to infectious	Allows for lag between exposure and illness; when agents can travel further so potentially be nearer more naïve population when infectious period starts.	1+	Median and mean both around 36 hrs	Random-normal distribution around population mean (36 h)

An estimated 50% of Americans [30] endorse at least one health-linked conspiracy theory Up to 44% of populace in diverse countries believe the demonstrable falsehood that vaccines cause autism [31]. Such beliefs have exacerbated real life disease outbreaks and risk-taking behaviour [32, 33]. Reported prevalence of beliefs in specific health myths (with health protection implications) linked to specific conspiracy theories among British and Americans ranges from 9% to 37% [30, 34]. The tendencies to believe CTs or poor quality information are distinct personal qualities [35], but empirical [23, 33, 36-38] and theoretical [38-41] evidence suggests extremely similar ideological and psychological processes underpin tendencies to believe both CTs and fake news. For model purposes, we assumed that predisposition to believe in CTs could serve as proxy for our posited reject-est attribute. Each individual's reject-est attribute did not change during the outbreak.

Published data [25] suggest that on average, British adults believed in 38.88% of CTs (SD 0.15, normal distribution around the mean). Therefore, reject-est values were assigned to agents such that the population mean = 38.88% (SD= 15%), constrained to range from 0 to 100% inclusive. Importantly, small groups (n=25) agents were distributed semi-randomly in the agent world such that they were physically clustered near others with similar reject-est values, and those individuals were also mutual members of each agent's information bubble (see below and additional item S2).

2.5 Information Bubbles

The phenomena that people choose how and from whom they receive information has been termed "filter bubbles" [26]. These bubbles work to discourage alternative viewpoints. For each agent, we generated a unique list of contacts in their bubble. The contact list included all agents within six spatial steps; because of deliberate placement earlier, many of these near-by individuals had similar reject-est values (assigned within the same octile of reject-est values). To this bubble were added approximately 120 agents anywhere in the agent-world, in a ratio about 2:1 similar:not similar reject-est octiles.. The target was to achieve final bubbles with a mean 150 members (range 80-230) to conform with Dunbar numbers, which estimate the number of persons with whom we each have significant (to us) relationships [29]. To reflect real world filter bubbles and social networks, ours should have variable reciprocity [42-45] and homophily levels [46], but veering towards demonstrating more rather than less reciprocity and homophily. Homophily is important in real health behaviour; individuals respond more to health promotion interventions when they come from a person or network of similar-to-recipient persons [47, 48]

Distributions of reject-est values, homophily and reciprocity were checked to confirm that the bubbles achieved desired attributes.

Membership of one's information bubble did not change during the simulated outbreak. Information sharing was independently decided from opportunities for physical contact. Real world sharing equivalents are telephone calls, sharing on social media, sending texts or emails, conversations, etc.

2.6 Advice spread

Two simultaneous processes happen during each model time-step. Agents move in a random direction, potentially transmit disease, incubate, are ill or recover. At the same time, pieces of advice are introduced (or "injected") into the community. Each injected piece of advice is exposed to just one agent. Injected advice has a 50:50 chance of being good/bad in the (no-intervention) stage 2 modelling. This individual responds to the information, as well as chooses whether to share it (decided stochastically). If advice was shared, the exposed individual made a separate and independent decision whether to share it again to others, creating an information cascade that continued until exhausted. Each sequence of information sharing started and completed within a single time step (one hour).

2.7 Predisposition to share advice

Our model rules for advice–sharing were designed to make the rumour distribution patterns resemble observations in Vosoughi et al [22] (about Twitter cascades). A *cascade* is a series of tweets with a single origin; cascade length is the maximum number of retweets passing thru only unique tweeters. The default likelihood of sharing good advice was set to 3%, because only about 3% of cascades were both >1 tweet long and demonstrably true stories. Vosoughi et al. reported several other cascade properties that were used as model targets: that the maximum depth for any true story was 9; (vs. 19 for false stories); 85% of cascades had depth = 1 (only tweeted once and not retweeted at all); 2% had depth > 5. Bots retweeted equal numbers of true and false stories, but humans overwhelmingly favoured retweeting false stories. Other research had similar observations as Vosoughi et al. about cascade depths and likelihood of sharing on Twitter [49-52].

Only about 15% of Twitter stories were shared. Of the information shared on Twitter, 80% was untrue stories (false rumours were four times more likely to be shared than true stories). We assumed that sharing of false stories is more likely among those with a counter hegemonic bias = agents with a high reject-est value. The likelihood of sharing bad advice ("willshare" variable) was calculated in stage 2-3 models for individual agents using Eq.1 which was found empirically to yield desirable cascade properties:

(**Eq1**) willshare = (3% * 4) * (reject-est / (mean [global reject-est of all agents])

Eq.1 causes the likelihood of sharing bad advice to be inflated from (default when good advice) 3% to 12%, and then further adjusted by the agent's reject-est value relative to the population mean reject-est. The net effect was a model assumption that agents with relatively higher reject-est values (more likely to believe in conspiracy theories) were more likely to share bad advice stories. Most real people don't repeatedly share the same information (good or bad). We applied the next formula to reset the willshare propensity after each share:

(**Eq2**) willshare = willshare / (4 ^ [number of times already-shared this advice])

Equations 1-2 are not meant to be definitive for social network behaviour. We determined these equations empirically and use them because they consistently led to cascade sharing patterns that agreed reasonably well with real observations in Vosoughi, Roy [22]

The model represents sharing by any means, including spoken conversation, phone calls, texts, social media, online forum postings, etc. When an agent shares advice, they only reach a very small fraction of people in their bubble (2.5%), which percentage made the cascade patterns behave reasonably well with regard to our targets for depth and onward sharing. Sharing behaviour was also simplified such that all shares for each cascade finished within each model time step (1 unit = 1 hour). Most real Twitter cascades stop growing within 2 hours of initiation [51].

2.8 Daily injections (introductions) of relevant discussions or stories

We used data on real number of daily conversations [53], and search frequency about health matters [54, 55] to estimate how many relevant information injections should happen in the model (10.4 per hour); more details how this was estimated are in Additional Item S1. We ran multiple simulations to find an injection rate (of advice) that led to the desired target of 10.4 cascades/hour (or 166 per day, based on 16 waking hours).

2.9 Finishing Stage 2: Bad advice making an outbreak worse

Changes in taking precautions we denote as ΔTP (absolute change in percentage of the time that precautions were taken, in response to each piece of advice an agent is exposed to). One aspect of ΔTP is partly evidenced from prior studies, given an assumption that bad advice is usually framed more emotionally. People change their statements about intended behaviour in response to exposure to information; they change behaviour more after frequent exposure [32, 56, 57].

However, at least in laboratory settings, the magnitude of change-in-intentions does not depend on whether material is emotively framed [58-62]. Therefore, our model assumes that ΔTP is the same whether advice is good or bad.

The final stage 2 model needed to achieve a net increase of 40% in the r0 in response to circulating information (from 1.9 to 2.66). Although ΔTP was the same fixed value in stage 2 models (whether good or bad advice), because more bad than good advice circulates (4:1 ratio), any ΔTP above zero increases r0 and tends to change other metrics such as attack rate and peak prevalence. Therefore we tested multiple values of ΔTP over the range .01 to 0.22 (1-100 iterations) to find a value of ΔTP that consistently produced the target r0 (2.66). We then compared the average outputs from the stage 2 model (50+ iterations) with results when intervention strategies were applied to try to reduce the impact of bad advice on the outbreak (stage 3).

2.10 Stage 3: Intervention Strategies

Proposed strategies to fight fake news include:

1) Provide counter-information that is equally or better evidenced, or more persuasive [2, 26, 63-66]
2) Tax the advertising or tax the profits of products sold via misinformation [67]
3) Drown bad info with good information [67]
4) Regulate information [26], possibly impose civil or criminal liabilities [2] which could lead to explicit censorship [2, 26]
5) Revise financial models available to fake news disseminators (incentives) to stop encouraging production and sharing of false (or even just very salaciously written) stories over truth and accuracy [3, 22, 28, 66]
6) Labelling (reliability rating or counter-arguments provided) by news provider [2, 22, 26, 66]
7) Encourage individuals to actively strive to make their own filter bubbles more diverse [26]
8) ‘Immunise’, recipients to disregard fake news (education-based strategy) [68]

We don’t model effects of intervention strategy 1 because the results are predictable; eg., good advice as contagious as bad advice is what happens in our stage 1 scenario (no net changes would result), and otherwise any changes will be linear responses if good advice increases without a reduction in bad advice. Pragmatically, we reduced strategies 2-8 to two basic interventions in stage 3 models, as described below. One hundred runs were tried for each intervention (tested separately rather than together), and the mean effects were reported and compared with each other and stage 1-2 outcomes. Stage 3 models were run under stage 2 conditions but with the below modifications:

- Reduce bad advice injections from 50% to 30% or 10% of total information exposures, to simulate tax disincentives, regulation, labelling or "drowning" strategies
- "Immunise" against bad information (but not against the virus, while able to react positively to good advice): a fixed percentage of randomly selected agents (30% or 90%) who never respond to or share bad advice, to simulate education-based or bubble-diversity strategies.

3. RESULTS

3.1 Model performance and optimisation exercises

With regard to information bubble construction, additional item S2 shows the spread of reject-est values, and that bubbles had high homophily and high reciprocity; bubble sizes also met Dunbar number targets. More details about the following results are in additional item S3. To achieve target r0 = 1.9, the optimal initialised mean take-precautions attribute for the models was 76.9%. At stage 2, we found that 138 advice injections per hour produced the target 166 conversations/day. This meant (over 20 iterations) that 70.7% of cascades had length = 1 (vs. target 85%) and about 1.83% of cascades had length ≥ 5 (vs. target 1.96%). We judged that the cascade results were acceptably close to targets. The stage 2 optimised ΔTP value was 0.026 (see model iterations in Item S3), which made r0 consistently rise from 1.9 to 2.66 in response to advice exposure.

3.2 Intervention strategies

Table 2 shows key outbreak indicators for the stage 1 model (no change in TP due to information spread) the final stage 2 model (with rate of advice injections = 138/hour and ΔTP = 0.026), and stage 3 models (what happens due to specific intervention strategies).

In Table 2, stage 2 is effectively a baseline to describe an outbreak exacerbated by circulating bad advice. Reducing the circulating bad advice from 50% to 30% of all introduced information, created a scenario that is much better than the stage 1 model, when circulating advice had no effect on average behaviour. Even if bad advice was reduced to 10% of total circulating information, the model still suggested that > 40% of individuals would get ill before the outbreak was finished.

'Immunising' 30% or more individuals (chosen at random, from any community bubble) tended to create an outbreak profile similar to or no worse than stage 1 (no influence of circulating information). This still meant almost 80% final attack rate and a peak prevalence near 24%. An immunisation rate of 90% produced r0 values around 1.38, with final attack rates over 50% and peak prevalence around 18%.

Additional item S3 shows a larger range of model assumptions and inputs than reported in Table 2, with respect to either altering the information balance or immunisation strategies. There was a clear trend towards more desirable outbreak measures (lower r0, lower final attack rate, lower peak prevalence) with less bad advice or higher immunisation rates.

4. DISCUSSION

With regard to reducing the amount of bad advice in circulation (whether by labelling poor quality info, drowning with better quality advice, regulation or financial disincentives), a reduction from 50% to 30% of total information exposures seems a large decrease but it may be feasible. Setting the ratio of good to bad advice to 70:30 more than negated the deleterious effects of circulating bad advice in our model. Even if 90% of the advice is good, however, some disease will still circulate (r0 stays above 1.0) because the baseline level of taking effective precautions is assumed to be imperfect (ie., well below 100%).

We were also interested in the 'herd immunity' levels required to 'immunise' people against fake news, and thus negate the influence of circulating bad advice on a hypothetical outbreak. The modelling suggests that any 'immunity' against bad advice reduces outbreak impacts. Herd immunity of at least 30% returned the outbreak to no worse than the stage 1 model scenario (ie, when circulating information has no impact).

Four previous studies used ABM to describe a norovirus outbreak [69-72], only one of which also incorporated information spread [69]. In other modelling, information spread led to increased awareness and greater protection against disease [73-82].

Similar to our study, some models [81-83] had behaviour outcomes comprised of multiple precautionary behaviours. Our clustering agent locations with respect to attitude towards trusting authority sources was novel, however. Considering how institutional distrust might change behaviour is also unusual in previous research [84].

Table 2. Stage 1 (no sharing), stage 2 (outbreak exacerbated by bad advice), and stage 3 (results using intervention strategies). Mean values for given outbreak characteristics, with 5-95th percentiles to indicate range without the most extreme values.

	r0	**Duration (days)**	**Final Attack Rate**	**Prevalence of illness at peak**	**# of iterations**
Stage 1					
No circulating advice	1.90	20.1	78.9%	23.5%	100
5-95th percentile range	1.80-1.99	15.2-25.9	76.0-81.4%	18.6-28.8%	
Stage 2. Circulating advice makes outbreak worse, r0 increase by 40%					
Good:Bad advice ratio is 50:50	2.66	19.0	91.8%	29.1%	
5-95th percentile range	2.50-2.89	15.1-25.1	90.3-93.8%	24.6-34.7%	100
Stage 3 models. strategies to reduce impacts of circulating bad advice in Stage 2 conditions					
Good:Bad advice ratio is 70:30	1.67	19.7	70.4%	21.2%	100
5-95th percentile range	1.53-1.78	15.4-26.3	63.6-75.5%	14.8-27.1%	
Good:Bad advice ratio is 90:10	1.22	14.1	41.5%	14.9%	100
5-95th percentile range	1.14-1.31	11.9-17.1	31.8-50.2%	10.1-19.8%	
30% of agents are ‘immunised’	1.91	20.2	79.0%	23.8%	100
5-95th percentile range	1.82-2.01	14.1-31.2	76.2-81.7%	18.2-28.7%	
90% of agents are ‘immunised’	1.38	17.1	53.8%	17.6%	100
5-95th percentile range	1.26-1.49	13.1-21.6	43.5-62.5%	11.0-23.3%	

Note: ‘immunised’ means immunity against believing or sharing bad advice, rather than immunity against norovirus.

4.1 Limitations

Limitations that prevent our results being fully generalizable to the real world are too many to fully list, we only try to consider the most important and feasible areas for improvement. The model was only tested for norovirus. Better data about true precautionary behaviour and behaviour change are the parameters that would most improve the reliability of our model outputs. Bayesian responses might also better reflect real world behaviour changes, too.

This model inherently considers the case of *Bad Advice*, presumed to be bundled with counter-hegemonic bias in contrast to *Good Advice* that is delivered with implied authority of endorsement from conventional sources. Bad advice that circulates for other reasons (well-meaning or dully presented but still incorrect) or good advice presented to be as exciting and 'contagious' as fake news [65] -- these are not included. Their omissions should only matter if the missing types of advice were thought to significantly modify the impacts of 'good' and 'bad' advice as described here.

The model also assumes that advice cascades terminate within a single hour; real information may spread over much longer time periods [22]. No agent is treated as more influential than others; there is inconclusive evidence about the importance of "influencers" in social networks [49, 65].

The model considers community, non-institutional settings (so not hospitals or parties or other high-density settings). No physical travel by new agents or existing agents to outside the system is considered. No adjustment was made for secretor status or innate immunity [85]. The models have a simplistic perspective on aspects of message framing. Framing and contextual presentation can be much more nuanced [86] in how they impact behaviour and beliefs. The only transmission pathway considered is person-to-person. In reality, many norovirus cases are contracted via fomites or food [87]. The model ignores the possibility of shedding before or after illness, which strongly raises r0 in norovirus outbreaks [85, 88]. There was no accounting for variations in immune response or age; infants and children are often more susceptible and have longer shedding periods. [85, 88]. We omitted foodborne, environmental or outside-illness shedding transmission pathways because they would have added extra complexity without adding extra clarity about how information sharing could affect outbreak development.

Agents 'immunised' against bad advice were randomly placed among the population, regardless of their reject-est attribute or local community traits; this is too simplistic and not realistic. Perhaps a 'vaccination' strategy analogous to ring vaccination or otherwise targeting demographic groups most likely to be susceptible to fake news would be more appropriate, when trying to 'immunise' people against fake news.

5. CONCLUSIONS

In our modelling, changing the ratio of good to bad advice (from 50:50 to 70:30) or at least 30% of people immunised to resist misinformation were both adequate thresholds to counteract negative impacts from fake news spreading during a norovirus outbreak. Changing the ratio of good:bad advice to 90:10 or immunising 90% of the population against misinformation was still not adequate to completely resist the impacts of circulating bad advice. How feasible it is to achieve these types of targets within communication networks or among community populations should be explored, with regard to cost-benefits and practical implementation.

REFERENCES

1. Tandoc EC, Lim ZW, Ling R. Defining "fake news" A typology of scholarly definitions. Digital Journalism. 2018;6(2):137-53.
2. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science. 2018;359(6380):1094-6.
3. Ball J. Post-truth: How bullshit conquered the world: Biteback Publishing; 2017.
4. Kim KA, Park J. Why do price limits exist in stock markets? A manipulation-based explanation. Europ Finan Manage. 2010;16(2):296-318.
5. Kosfeld M. Rumours and markets. J Math Econ. 2005;41(6):646-64.
6. Persily N. Can democracy survive the Internet? Journal of Democracy. 2017;28(2):63-76.
7. Speed E, Mannion R. The rise of post-truth populism in pluralist liberal democracies: challenges for health policy. International Journal of Health Policy and Management. 2017;6(5):249.
8. Willem L, Verelst F, Bilcke J, Hens N, Beutels P. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). BMC Infect Dis. 2017;17(1):612.
9. CDC. Norovirus: Clinical Overview 2016 [updated Feb 21 2013. Available from: https://www.cdc.gov/norovirus/hcp/clinical-overview.html.
10. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut. 2011:gut. 2011.238386.
11. Knapton S. NHS 111 calls peaked in week before Christmas amid rising cases of flu and norovirus. The Telegraph. 2017 29 December.
12. Serrano A. New Strain Of Norovirus Hits Hard. CBS Evening News. 2007 1 April.
13. Snug I. Norovirus caused the equivalent of two entire hospitals to be closed this winter. Nursing Notes. 2018 8 March.
14. Staff. Closing the door on winter vomiting bug. The Irish Times. 2007 16 January.
15. Staff. Hospitals overwhelmed by flu and norovirus patients. CTVNewsca. 2013 10 January.
16. Yang C, Wilensky U. Netlogo epidem basic model. Center for Connected Learning and ComputerBased Modeling Northwestern University Evanston IL, pages Evanston, IL. 2011.
17. Fretz R, Svoboda P, Lüthi T, Tanner M, Baumgartner A. Outbreaks of gastroenteritis due to infections with Norovirus in Switzerland, 2001–2003. Epidemiol Infect. 2005;133(3):429-37.
18. Lee RM, Lessler J, Lee RA, Rudolph KE, Reich NG, Perl TM, et al. Incubation periods of viral gastroenteritis: a systematic review. BMC Infect Dis. 2013;13(1):446.
19. Lopman B, Reacher MH, Vipond IB, Sarangi J, Brown DW. Clinical manifestation of norovirus gastroenteritis in health care settings. Clin Infect Dis. 2004;39(3):318-24.

20. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74.
21. Gaythorpe K, Trotter CL, Lopman B, Steele M, Conlan A. Norovirus transmission dynamics: a modelling review. Epidemiol Infect. 2018;146(2):147-58.
22. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359(6380):1146-51.
23. Narayanan V, Barash V, Kelly J, Kollanyi B, Neudert L-M, Howard PN. Polarization, Partisanship and Junk News Consumption over Social Media in the US. arXiv preprint arXiv:180301845. 2018.
24. Sunstein CR, Vermeule A. Conspiracy theories: Causes and cures. J Pol Phil. 2009;17(2):202-27.
25. Brotherton R, French CC, Pickering AD. Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. Front Psychol. 2013;4:279.
26. Garrett RK. The “echo chamber” distraction: Disinformation campaigns are the problem, not audience fragmentation. Journal of Applied Research in Memory and Cognition. 2017;6(4):370-6.
27. Radu RN, Schultz T. Conspiracy Theories and (the) Media (Studies). SSRN. 2017:9.
28. Shane S. From Headline to Photograph, a Fake News Masterpiece. The New York Times. 2017 Jan 18.
29. Dunbar RI, Arnaboldi V, Conti M, Passarella A. The structure of online social networks mirrors those in the offline world. Social Networks. 2015;43:39-47.
30. Oliver JE, Wood T. Medical conspiracy theories and health behaviors in the United States. JAMA Internal Medicine. 2014;174(5):817-8.
31. Ipsos. Perils of Perception. 2017.
32. Gorski D. Outbreaks among Somali immigrants in Minnesota: Thanks for the measles again, Andy: Science-Based Medicine; 2017 [Available from: https://sciencebasedmedicine.org/outbreaks-among-somali-immigrants-in-minnesota-thanks-for-the-measles-again-andy/.
33. Swami V, Voracek M, Stieger S, Tran US, Furnham A. Analytic thinking reduces belief in conspiracy theories. Cognition. 2014;133(3):572-85.
34. Democrats and Republicans differ on conspiracy theory beliefs [press release]. 2013.
35. Surma K, Oliver E. Believe It or Not? Credulity, Skepticism, and Misinformation in the American Public. American Politics Workshop; March 5; Madison, Wisconsin2018. p. 41.
36. Uscinski JE, Klofstad C, Atkinson MD. What drives conspiratorial beliefs? The role of informational cues and predispositions. Pol Res Quarterly. 2016;69(1):57-71.
37. Dagnall N, Drinkwater K, Parker A, Denovan A, Parton M. Conspiracy theory and cognitive style: a worldview. Front Psychol. 2015;6:206.
38. Pennycook G, Rand DG. Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. SSRN. 2017:54.
39. Kahan DM. Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. SSRN. 2017:9.
40. Allan M. Information Literacy and Confirmation Bias: You can lead a person to information, but can you make him think? 2017.
41. Moulding R, Nix-Carnell S, Schnabel A, Nedeljkovic M, Burnside EE, Lentini AF, et al. Better the devil you know than a world you don't? Intolerance of uncertainty and worldview explanations for belief in conspiracy theories. Pers Individ Dif. 2016;98:345-54.
42. Jiang B, Zhang Z-L, Towsley D, editors. Reciprocity in social networks with capacity constraints. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015: ACM.
43. Kwak H, Lee C, Park H, Moon S, editors. What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web; 2010: ACM.
44. Lerman K, Ghosh R. Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. Icwsm. 2010;10:90-7.

45. Zhu Y-X, Zhang X-G, Sun G-Q, Tang M, Zhou T, Zhang Z-K. Influence of reciprocal links in social networks. PLoS One. 2014;9(7):e103007.
46. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences. 2009;106(51):21544-9.
47. Barclay KJ, Edling C, Rydgren J. Peer clustering of exercise and eating behaviours among young adults in Sweden: a cross-sectional study of egocentric network data. BMC Public Health. 2013;13(1):784.
48. Centola D, van de Rijt A. Choosing your network: Social preferences in an online health community. Soc Sci Med. 2015;125:19-31.
49. Bakshy E, Hofman JM, Mason WA, Watts DJ, editors. Everyone's an influencer: quantifying influence on twitter. Proceedings of the fourth ACM international conference on Web search and data mining; 2011: ACM.
50. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res. 2015;17(6).
51. Lerman K, Ghosh R, Surachawala T. Social contagion: An empirical study of information spread on Digg and Twitter follower graphs. arXiv preprint arXiv:12023162. 2012.
52. Zubiaga A, Liakata M, Procter R, Hoi GWS, Tolmie P. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS One. 2016;11(3):e0150989.
53. Morning Advertiser. Pub is hub of conversation 2010 [Available from: https://www.morningadvertiser.co.uk/Article/2010/08/18/Pub-is-hub-of-conversation.
54. Lau T, Horvitz E. Patterns of search: analyzing and modeling web query refinement. UM99 User Modeling: Springer; 1999. p. 119-28.
55. Pew Internet Project. Health Online 2013. Pew Research Center, California Healthcare Foundation; 2013 January 15.
56. Branswell H. Measles sweeps an immigrant community targeted by anti-vaccine activists: STAT; 2017 [updated 8 May 2017. Available from: https://www.statnews.com/2017/05/08/measles-vaccines-somali/.
57. Noar SM. A 10-year retrospective of research in health mass media campaigns: Where do we go from here? Journal of Health Communication. 2006;11(1):21-42.
58. O'keefe DJ, Jensen JD. The advantages of compliance or the disadvantages of noncompliance? A meta-analytic review of the relative persuasive effectiveness of gain-framed and loss-framed messages. Annals of the International Communication Association. 2006;30(1):1-43.
59. Arora R, Stoner C, Arora A. Using framing and credibility to incorporate exercise and fitness in individuals' lifestyle. Journal of Consumer Marketing. 2006;23(4):199-207.
60. Detweiler JB, Bedell BT, Salovey P, Pronin E, Rothman AJ. Message framing and sunscreen use: gain-framed messages motivate beach-goers. Health Psychol. 1999;18(2):189.
61. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med. 2011;43(1):101-16.
62. Lithopoulos A, Bassett-Gunter RL, Martin Ginis KA, Latimer-Cheung AE. The effects of gain-versus loss-framed messages following health risk information on physical activity in individuals with multiple sclerosis. Journal of Health Communication. 2017;22(6):523-31.
63. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? Br Med J. 2014;349:g6178.
64. Haer T, Botzen WW, Aerts JC. The effectiveness of flood risk communication strategies and the influence of social networks—Insights from an agent-based model. Environ Sci Policy. 2016;60:44-52.
65. Berger J, Milkman KL. What makes online content viral? J Marketing Res. 2012;49(2):192-205.

66. Lyons T. Replacing Disputed Flags with Related Articles: Facebook; 2017 [updated 20 December. Available from: https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/.
67. Glaeser EL, Ujhelyi G. Regulating misinformation. J Public Econ. 2010;94(3-4):247-57.
68. Rochlin N. Fake news: belief in post-truth. Library Hi Tech. 2017;35(3):386-92.
69. Hill AL, editor Norovirus outbreaks: Using agent-based modeling to evaluate school policies. Proceedings of the 2016 Winter Simulation Conference; 2016: IEEE Press.
70. Bartsch SM, Huang SS, Wong KF, Avery TR, Lee BY. The spread and control of norovirus outbreaks among hospitals in a region: a simulation model. Open Forum Infectious Diseases. 2014;1(2).
71. Yu B, Wang J, McGowan M, Vaidyanathan G, editors. Agent-based stochastic simulations of shipboard disease outbreaks. Proceedings of the 2010 Spring Simulation Multiconference; 2010; Orlando FL: Society for Computer Simulation International.
72. Gutierrez LM. Agent-based simulation of disease spread aboard ship. Monterey CA: Naval Postgraduate School; 2005.
73. Bisset KR, Feng X, Marathe M, Yardi S, editors. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. Winter Simulation Conference; 2009: Winter Simulation Conference.
74. d'Onofrio A, Manfredi P. Information-related changes in contact patterns may trigger oscillations in the endemic prevalence of infectious diseases. J Theor Biol. 2009;256(3):473-8.
75. Durham DP, Casman EA. Incorporating individual health-protective decisions into disease transmission models: a mathematical framework. J Royal Soc Interf. 2011:rsif20110325.
76. Mao L. Predicting self-initiated preventive behavior against epidemics with an agent-based relative agreement model. Journal of Artificial Societies and Social Simulation. 2015;18(4):6.
77. Smith MC, Broniatowski DA, editors. Modeling Influenza by Modulating Flu Awareness2016; Cham: Springer International Publishing.
78. Zhang H-F, Xie J-R, Chen H-S, Liu C, Small M. Impact of asymptomatic infection on coupled disease-behavior dynamics in complex networks. EPL (Europhysics Letters). 2016;114(3):38004.
79. Hatzopoulos V, Taylor M, Simon PL, Kiss IZ. Multiple sources and routes of information transmission: implications for epidemic dynamics. Math Biosci. 2011;231(2):197-209.
80. Chen F, Jiang M, Rabidoux S, Robinson S. Public avoidance and epidemics: insights from an economic model. J Theor Biol. 2011;278(1):107-19.
81. Andrews MA, Bauch CT. Disease interventions can interfere with one another through disease-behaviour interactions. PLoS Comput Biol. 2015;11(6):e1004291.
82. Guo D, Li KC, Peters TR, Snively BM, Poehling KA, Zhou X. Multi-scale modeling for the transmission of influenza and the evaluation of interventions toward it. Sci Rep. 2015;5:8980.
83. Fast SM, González MC, Wilson JM, Markuzon N. Modelling the propagation of social response during a disease outbreak. J Royal Soc Interf. 2015;12(104):20141105.
84. Fajebe A. Computational modeling of spontaneous behavior changes and infectious disease spread [PhD]: Georgia Institute of Technology; 2016.
85. Simmons K, Gambhir M, Leon J, Lopman B. Duration of immunity to norovirus gastroenteritis. Emerg Infect Dis. 2013;19(8):1260.
86. Holton A, Lee N, Coleman R. Commenting on health: A framing analysis of user comments in response to health articles online. Journal of Health Communication. 2014;19(7):825-37.
87. Lopman B, Gastanaduy P, Park GW, Hall AJ, Parashar UD, Vinjé J. Environmental transmission of norovirus gastroenteritis. Curr Opin Virol. 2012;2(1):96-102.
88. Milbrath M, Spicknall I, Zelner J, Moe C, Eisenberg J. Heterogeneity in norovirus shedding duration affects community risk. Epidemiol Infect. 2013;141(8):1572-84.