UNIVERSITÀ DEGLI STUDI DI TRENTO

DEPARTMENT OF MATHEMATICS

Doctoral School in Mathematics
CYCLE XXXI

SOCIAL DYNAMICS AND BEHAVIORAL RESPONSE DURING HEALTH THREATS

PAOLO BOSETTI

Advisors:

Andrea Pugliese
University of Trento, Trento

Stefano Merler
Fondazione Bruno Kessler , Trento

Bruno Lepri
Fondazione Bruno Kessler , Trento

ACADEMIC YEAR 2018/2019

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

The spread of infectious diseases in human populations has always represented one of the major threats to human existence [6, 117, 79, 110]. In fact, dealing with communicable diseases remains one of the greatest challenges for public health systems across the globe [160, 120, 109, 81].
To tackle this challenge, the study of infectious diseases evolved in a very interdisciplinary field, driven by the necessity of simultaneously considering the wide variety of factors that are involved [17]. Disciplines such as classical epidemiology, mathematics, biology, and social science converge their focuses to shed light on the complex dynamics underlying the spread of the infections [74].

Several different factors, such as the pathogen's biology, the nature of the disease vectors, and the chances of interactions between species, heavily shape the emergence and diffusion of infectious diseases [6, 82, 40, 94]. As long as humans are concerned, the spreading and unfolding of infections can be deeply affected by human behavior and human choices, as well as by the structural and dynamical characteristics of our society [46, 98, 59, 57, 155, 161]. On one hand, human behavior and human choices imply that individuals are not passive agents but they may react in response to a perceived risk or threat [121, 122, 119, 80, 45]. On the other hand, since individuals are part of the society, social norms and social dynamics guide and influence human interactions [74, 8, 87, 130].

Mathematical models are among the most powerful tools to deal with such complexity, capable to quantitatively characterize the emergence of infection and to predict the development and evolution of the outbreak. Their effectiveness benefits from the technological revolution and the increasing dialogues with different research fields.
Classical mathematical models of infectious diseases consider all the individuals inside the population as identical and with a static behavior over time [6]. This assumption is related to the fact that, in the past decades, data on individual's behaviors were difficult to obtain and usually demanded a significant effort in their collection and analysis. Information extracted through traditional means (e.g. surveys and interviews) are not usually continuous over time and they are not able to precisely assess behavioral changes, even when collected for a sufficiently long time interval.

The digital revolution, that we are experiencing nowadays, delivers an unprecedented quantity of high frequency human behavioral data [131]. This paves the way to new opportunities in modeling infectious diseases dynamics and allows to unravel hidden patterns and describe society as a whole [87, 156, 116, 87]. For the first time,

we have the possibility to include individual and spatially fine grained behavioral patterns in the models. Such *data driven* models are capable of detecting the major determinants in the spread of an epidemic, to asses a possible change in the behavior in response to a threat and to suggest effective policies that help to prevent the spread of diseases [156].

Two relevant aspects involving the dynamics of infectious diseases associated with human behaviors can be distinguished:
i) Behavioral changes in response to pandemic threats;
ii) Impact of the interactions between groups, i.e. social dynamics of the population that might boost and sustain the spread of the infections.

1.1 HUMAN BEHAVIOR IN RESPONSE TO PANDEMICS

As previously mentioned, behavioral response triggered by epidemic threats can play a key role in shaping the transmission dynamics of infectious diseases. Risk perception may affect the acceptability of public health policies as well as induce adaptive behavioral changes to reduce the individual risk of infection [46, 59, 45, 98, 123, 145, 141].
Many examples of adaptive responses to a pandemic are documented in the literature [71, 112, 121, 168]. For instance, during the Severe Acute Respiratory Syndrome (SARS) epidemic in 2003, individuals protected themselves by wearing masks, avoiding public transportation and improving personal hygiene measures. Similar adaptive behavior led by increasing social distance was reported and investigated also for pandemic flu of the 1918 and for the H1N1 2009 pandemic [23, 121, 92, 145]. Remarkably, these spontaneous reactive behavioral changes might represent a significant protective factor during the initial stage of the epidemic by slowing down the exponential growth of the infections and increasing the chances of a successfully containment of the disease [128, 37].

From a different perspective, behavioral changes can be driven by the adoption of containment policies implemented by the governments to prevent widespread diseases [47, 48, 170, 27, 9]. When outbreaks occur, policy makers must implement strategies aiming to interrupt the transmission chain of the infections. During the West Africa Ebola outbreak in 2014, several control measures were adopted [115, 157, 107]. These included social distancing interventions, such as school closures, annulment of mass gathering events, quarantine of individuals exposed to the disease, safe burials, and borders restrictions. Efforts towards adopting public health measure were rewarding, at least to limit the epidemic strength. However, a number of studies have evidenced that restrictive measures can have severe economic and social impli-

cations, being detrimental for the country experiencing the outbreak [127, 99, 118, 135, 20]. Ineffective communications between policy makers and citizens may jeopardize the efficacy of the measures and reduce their acceptance, triggering negative dynamics and amplifying social distress. Conversely, community participation coordinated by health care systems are associated with successful outcomes in reducing the burden of disease [2, 44, 4].

It is a matter of fact that behavioral changes in the populations depend on the *risk perception* and individual's *awareness* about the imminence of a health threat. As such, understanding the role played by the awareness and its associated drivers represents a crucial aspect to model the dynamics of behavioral changes in the population. In particular, higher risk perception driven by the awareness fosters reactive responses of both individuals and government, in order to take precautionary actions. Moreover, local spread of awareness in the community experiencing the epidemic can control the diffusion of the disease [58, 56] and can be exploited in order to contain and mitigate the outbreak [70, 93].

Media communications increase the individuals' awareness and coordinate cooperations between policy makers and citizens [159]. In fact, many examples evidenced that risk perception does not mimic the disease prevalence because of several distortion factors, such as the misrepresentation of the health threat by news media or peer personal opinions [98, 59, 155, 57, 55, 143, 145]. This fact poses a challenge in the inclusion of awareness in operational forecasting frameworks.

The concept of individuals' awareness is usually nested in mathematical models of individuals behavioral response. In particular, a lot of works couple the awareness with the prevalence of infections inside a population [59, 155]. Because of the aforementioned distortion factors, this approach might not be effective. One of the main problems is that the role played by media drivers and peer influence in the building of awareness lacks a quantitative characterization, mostly because data describing the information level of an individual during the outbreaks are difficult to obtain by traditional means. Thus, considerable effort is needed to understand the different drivers of information about health threats to include awareness directly in the epidemic models to ultimately design optimized information campaign and increase the efficacy of both spontaneous behavioral changes and mitigation policies.

1.2 SOCIAL DYNAMICS AND SPREAD OF INFECTIOUS DISEASES

Human-to-human transmissions of an infectious disease are led by social factors that guide the contacts between individuals [46, 74, 15, 132, 8]. Investigating the social dynamics in the population, such as the heterogeneity of contacts and the structure of social groups, reinforce reliable predictions of the future evolution of an epi-

demic [161]. As an example of the primary role of heterogeneity of contacts in the population, one can consider a few individuals with high degrees of contact who infect the majority of the populations. In opposition of considering all the individuals as identical, the role of super-spreaders is crucial in studying the dynamics of the infections, as widely documented in the literature [95, 60, 169, 85]. An other aspect is the composition and social interactions between groups. A typical human behavior that people exhibit is homophily [102]. This refers to the tendency of people to cluster together in groups that share common characteristics, such as age, belief, ethnicity etc. When considering an epidemic that diffuses across the homophylic clusters, the knowledge of the community structure and dynamics represents a relevant tool to model the spread of an infectious disease. Models that account for these types of characteristics can both evaluate the contribution of each group and, in turn, identify the adequate strategies in the struggle against the disease [77, 75, 113, 72, 7]. Accordingly, assumptions of well mixed population might mask the role of groups in originating outbreaks and potentially sustaining the epidemics. A case in point is represented by the measles outbreak in the Amish communities of US in 2014 [61, 62]. The Amish are characterized by sub-optimal coverage level for the most basic vaccines, mostly due to their rejection of modern technologies and medicine. In fact, vaccination is not forbidden by the Amish religion, but personal beliefs shared among the communities result in low participation in the preventive health care system of the country and in a critical level of vaccines. This critical behavior results in immunity gaps and it is one of the factors associated with the re-emergence of measles in the United States, with hundreds of cases inside the Amish communities.

Social dynamics and group structures are not a static characteristic of the population. Disruptive events such as economic or political crises can rapidly alter the group composition and interactions in the system. In particular, such events can lead to massive migrations which heavily affect the social structure both of origin and arrival countries [50, 153]. From 2010 to 2017 global migrations have increased of almost 50%, going from 173 to 258 million people leaving their countries, with a trend that is expected to grow in the future [3]. This poses a challenge for infectious diseases modelers in forecasting and capturing the evolution of demographic changes and interactions in the populations [3, 173].

In particular, many developed countries, that experience the arrivals of migration flows, often perceive refugees as a threat to their economic and health system [31]. This happens despite the documented evidences that hosting counties benefit from immigration: refugees are typically younger and healthier than the local population and they can actively contribute to the economic growth [3]. However, this biased perception on migrants might ignite mechanism of segregation that is detrimental for both populations and potentially boosts, rather than hamper, the spreading of

infectious diseases. That is why understanding migration flows and their impact on society is of great importance to help and guide public health preventive measures.

1.3 HUMAN BEHAVIOR AND EPIDEMIC SPREADING IN THE DIGITAL ERA

It is a matter of fact that we live in a connected world. Everyday people generate, share, and acquire a huge amount of information thanks to technological progress. As a consequence of the ubiquity of mobile devices, we are now facing an explosive growth of digital traces left behind by individuals [131]. Data are stored, computed and analyzed to extract prevalent patterns capturing a broad spectrum of human behaviors [87]. Undoubtedly, the digital age has opened the path to new opportunities for researchers, giving the chance to access phenomena and measures once invisible. Remarkably, this represents a great opportunity for developing countries that usually lack of official statistics and infrastructures to collect it [21].
For example, in 2015 Blumenstock and colleagues combined different sources of data, i.e. a small set of data collected from surveys and data on the mobile activity of 1.5 Million of customers, to build a high resolution map of poverty of the districts of Rwanda [22]. The strength of Blumenstock's work was the capability of giving a snapshot of the wealth of the country ten times faster and fifty times cheaper than required by the classical Demographic Health Survey.

Thanks to the increasing penetration rate of mobile phones, mobile data represent an increasing opportunity to assess behavioral and mobility pattern and complement traditional resources, such as surveys and interviews [131, 19, 21]. Information about mobile communications is usually collected for billing purposes by telecommunication companies and stored in form of Call Detailed Record (CDR). Each CDR usually contains metadata about the caller and callee and information about the location and time of the call; frequently also demographic information about the costumer are included. These data have been extensively used to build realistic fine grained models of human mobility [14, 10, 12, 66]. Such human mobility models can be used to map the mobility flows between geographical areas at different scales and to improve modeling of spatial spreading of infectious diseases[10, 11, 104, 9, 163, 78, 93].

On a different but related note, data on social media and online platforms are used to asses the risk perception on a health threat and also to forecast the evolution of an epidemic. For example, textual information about the tweets on Zika and Ebola outbreaks helped to measure the level of anxiety in the populations related to the fear of the epidemics [55, 143, 150]. Specifically, tweets associated with negative emotions were higher compared with those associated with recurrent outbreak such as seasonal influenza. Quantifying human perceptions on public emergence can be cru-

cial to avoid misperception about a pandemic and, in turn, design optimized awareness campaigns [84]. Additionally, data obtained from the number of Google searches and the visits to a Wikipedia page can produce a timely and accurate assessment of the trend of the epidemic [134, 133, 101]. This original and valuable approaches can be coupled to existing sentinel surveillance to have real time nowcasting of the epidemics [133, 63, 88]. What so far exposed represents only a few examples of the great potential that the digital revolution has and dealing with this kind of data will be a great aid for the future of infectious disease modeling.

1.4 OUTLINE OF THIS THESIS

This dissertation argues about two different aspects of human behavior involving the dynamics of infectious diseases, as discussed in the previous sections. The first one deals with the development and analysis of epidemiological models aiming at better understanding the spread of awareness in human populations during epidemic threats. The second one consists of a data driven computational model to assess the potential risk of experiencing measles re-emergence in Turkey as a consequence of the great concentration in the country of Syrian refugees not adequately immunized against measles.

One of the most original aspects of this dissertation is represented by the usage of innovative digital data made recently available. These include data obtained from the use of mobile phones to infer human mobility patterns among refugees, and digital records on the individual use of the Internet coupled with mechanistic transmission model to assess the drivers of awareness about new epidemic threats.

In chapter 2 we propose a framework able to disentangle and characterize the contribution of the two drivers in the building of awareness of individuals about infectious diseases. In particular, the modeling approach is tested for several major recent epidemics, such as Ebola and Zika, by accurately fitting a proxy of the level of awareness in the population; i.e. the time series of Wikipedia page views related to the disease originating the epidemic. Obtained results consistently suggest that both word of mouth and media communications represent crucial drivers of the dynamics of individual' concern about the considered epidemics. The proposed analysis represents the first attempt to identify the major determinants of the spread of awareness, grounding a modeling approach on the availability of digital data on individual use of the Internet

In chapter 3 We assess the potential risk of experiencing measles re-emergence in Turkey as a consequence of the great concentration in the country of Syrian refugees. In particular, Turkey hosts almost 3.5M refugees and has to face a humanitarian emer-

gency of unprecedented levels. We use mobile phone data to map the mobility patterns of both Turkish and Syrian refugees, and use these patterns to build data-driven computational models for quantifying the risk of epidemics spreading for measles, a disease having a satisfactory immunization coverage in Turkey but not in Syria, due to the recent civil war, while accounting for hypothetical policies to integrate the refugees with the Turkish population. Our results provide quantitative evidence that policies to enhance social integration between refugees and the hosting population would reduce the transmission potential of measles, preventing the onset of widespread large epidemics in the country. Our results suggest that social segregation does not inhibit but rather boosts potential outbreaks of measles to a greater extent in Syrian refugees but also in Turkish citizens, although to a lesser extent.

In chapter 4 presents a summary and discussion of the main results of the dissertation

2

DISENTANGLING SOCIAL CONTAGION AND MEDIA DRIVERS IN THE EMERGENCE OF HEALTH THREATS AWARENESS

2.1 INTRODUCTION

In recent years, thanks to the increasing abundance of (near) real-time, high-quality data on populations, human mobility patterns, and pathogens' biology, the use of data-driven computational models for the study of epidemic outbreak response has gained considerable traction in the public health community [136]. In this context, predictive epidemic modeling is emerging as an inter-disciplinary field that has been recently used to support responses to recent outbreaks such as the 2009 H1N1 pandemic, the 2014 West Africa Ebola outbreak, the Zika epidemic in the Americas in 2016, and the Highly Pathogenic Avian influenza A(H7N9) in 2014 and 2015.

In view of the huge potential of predictive modeling, we must also be aware of the challenge inherent to the real-time modeling of the feedback loop between the disease progression and the response of social systems. Recent examples (Ebola, MERS, etc. [157, 125]) have shown that the spreading of infectious diseases strongly depends on the social adaptive behavior that characterizes the reaction of the population to the awareness and the perceived risk in the face of the epidemic [157, 125, 46]. In other words, the predictive power of mathematical and computational models heavily relies on our understanding of the population awareness to the disease and the ensuing behavioral changes such as social distancing, travel limitations, etc. [145, 48, 33, 121].

Several mathematical and computational models of the feedback mechanisms between diseases spreading and the effects due to the awareness of the epidemic in the population has been put forward in the modeling community [121, 58, 119, 123, 150, 45]. In particular, two main mechanisms have been invoked in the spreading of awareness and fear of epidemics: i) the massive flow of news from mass media that possibly acts as an exogenous force on the global population [150, 55, 146]; and ii) the social contagion due to the word of mouth and peer influence [121, 58, 119, 123]. However, little is known about the relative contribution of the two mechanisms in shaping the spread of information in the population [86, 150, 139], and to what extent individuals' response is affected by mechanisms of social reinforcement [28]. As a matter of fact, the richness of models and mathematical approaches, has not yet

been transferred to any operational forecasting framework; mostly due to the lack of a quantitative data-driven characterization of the different mechanisms affecting the spreading of the awareness in the population [150, 42].

Here, we propose a modeling framework able to disentangle the contribution due to media drivers and the social contagion in the awareness building of infectious diseases. The developed model assumes that individuals become aware of an epidemic either by means of media communications on different stories related to the ongoing epidemic (through newspapers, websites, broadcasts, etc.) or as a consequence of conversations with other individuals. We test our modeling approach for several major recent epidemics by accurately fitting a proxy of the level of awareness in the population; i.e. the time series of Wikipedia page views [51, 52] related to the disease originating the epidemic. The underlying assumption here is that among the individuals that develop awareness, a fraction them initiate an information seeking behavior using Wikipedia to address this need. More precisely, we fit the proxy signal by using a non-linear time dependent mechanistic model explicitly accounting for the combination of temporal changes in the media volume, as recorded in the Google News platform [68] and treated as an exogenous factor of the system, and an endogenous contagion process driven by risk perception and spontaneous social interactions between individuals. The proposed non-linear model is able to measure the effect of the endogenous progression of awareness in the population and quantify its contribution relative to the news and the media drivers. We perform a thorough information theoretical model selection analysis [24] showing that the proposed model outperforms supervised machine learning approaches based only on the news volume, and non-linear models accounting only for the endogenous social contagion component. Our approach outperforms the other considered models also in out-of-sample predictive tests. Interestingly, the mechanistic modeling proposed here allows the estimate of specific parameters such as the doubling time and transmissibility of the awareness through the different routes of information considered. Thus, this work has the potential to open the path to the inclusion of the diffusion of awareness in a wide range of biological and social contagion models, allowing the measurement of specific parameters of social contagion and the design of optimized awareness campaigns.

2.2 RESULTS AND DISCUSSION

We consider as a proxy of the awareness of major infectious disease epidemics the time series of Wikipedia page views for several health threats in the US and Italy. In particular, the 2014 Ebola Outbreak in West Africa [157, 65] and the spread of Zika virus in the Americas during 2015 [172] are considered illustrative case studies to in-

vestigate the impact of *Public Health Emergencies of International Concerns* on the individuals' awareness in both US and Italian populations. Meningitis in Italy and Influenza in US between 2016 and 2017 are considered in order to investigate, respectively, the effect of local and persisting epidemiological threats [55, 147].

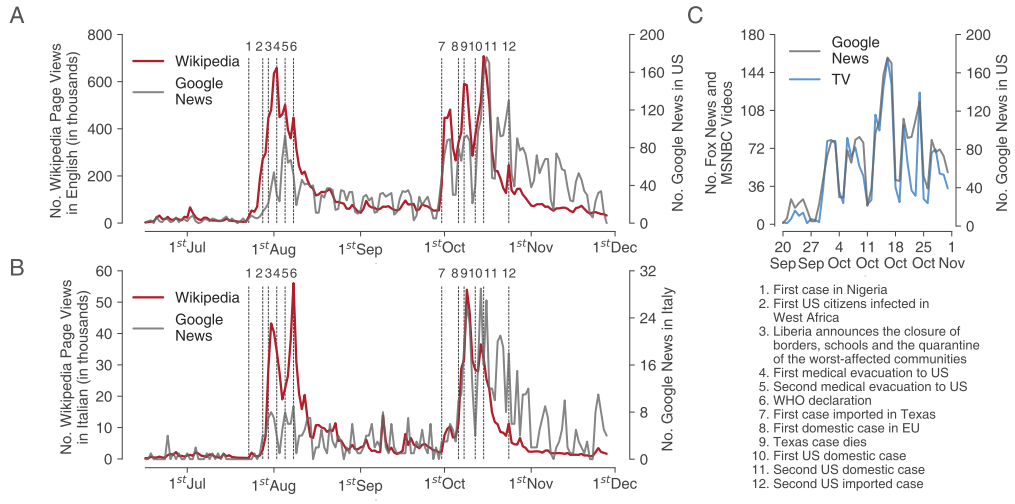

Figure 1: Proxies of health threat awareness and media attention for the illustrative case of Ebola epidemic. **A)** Red line represents the daily number of page views on Wikipedia articles related to Ebola infection during 2014 from the English version of Wikipedia [51]. Grey line represents the daily number of news released on Ebola in the US, as obtained from the Google News platform [68]. Dotted lines indicate noticeable events associated with the West Africa Ebola epidemic. **B)** As A) but for the Italian version of Wikipedia and news released in Italy. **C)** Comparison of two different proxies of media attention to the Ebola epidemic. Grey line represents the daily number of news released on Ebola in the US, as obtained from the Google News platform [68]. Blue line indicates the number of Ebola related videos per day, from Fox News and MSNBC [150].

The level of media attention to the spreading of an infectious disease is certainly contributing to the population awareness of the disease. This can be readily observed by comparing the time series of the Wikipedia accesses and the volume of news on the epidemic measured from the Google News platform, a quantitative proxy of the media attention over time. As shown in Fig. 1 for the exemplary cases of the 2014 Ebola, the Wikipedia accesses and news volume time series have an extremely good correlation when no or negligible time lag is considered between the two signals (see the Appendix for a detailed analysis). A closer inspection to the curves however shows that, generally, the awareness tends to increase and decrease at a faster rate before and after the peak of the news cycle, respectively. For instance, in the US, after a huge volume of Wikipedia accesses and news about Ebola during week 42 of 2014, Wikipedia page views dropped by 64% in the following week, while media coverage

decreased by 22% only, remaining significantly high up to mid-November. Similar temporal patterns were observed when comparing news coverage on Zika and individuals' online behavior in US (see Appendix), Guatemala and Brazil [143]. These features are the fingerprint of acceleration and saturation phenomena that can be ascribed to endogenous word of mouth processes. Here, as word of mouth processes we can consider all personal communications among individuals, including both real-world and on-line contacts.

In our model, we assume that the overall strength of media coverage in a given day is proportional to the overall amount of news in that day, which include the contribution of both more and less relevant communications. In support of such assumption, it is worth noting that major peaks in the Google News time series associated with Ebola follows the occurrence of some relevant events for the two considered countries, as the detection and death of the first Ebola case within the US borders, and the first transmission recorded in EU (Fig. 1 A,B). In order to test the representativeness and appropriateness of Google News data to model the media attention over time, we compared the temporal dynamics of the number of Ebola related videos per day from Fox News and MSNBC [150] with the corresponding Google News signal (Fig. 1 C). We found that the two signals are strongly correlated (Pearson correlation coefficient 0.92 p-value < 0.001) therefore suggesting that Google News represents a good proxy of the variation in strength of media coverage over time.

In order to model mechanistically the awareness spreading in the population we use a SIRS transmission mechanism [6]. In the transmission model susceptible individuals (S) represent people unconcerned or unaware about an epidemic threat that can get informed (I) either through the word of mouth, as a consequence of a social contagion based on peer-to-peer individual interactions with informed individuals (I), or directly from the media (M). Potential development of immunity (R) against the exposure to new information and possible mechanism of immunity waning is fully considered with the SIRS model. The model is calibrated by assuming that at each time stamp t a fraction k of individuals becoming aware of the disease seeks further information on Wikipedia, which is a proxy of awareness acquired through other means. At each time stamp the model keeps track of all the newly become aware individuals $\mathrm{I}_t^{\mathrm{New}}$; i.e. the number of individual transitions $\mathrm{S} \rightarrow \mathrm{I}$. The Wikipedia page views (W_t) at time t is then given by the following relation

$$\mathrm{W}_t = k\mathrm{I}_t^{\mathrm{New}} + \mathrm{W}_0\,,$$

where W_0 represents the number of Wikipedia page views in absence of media attention. The number of individuals $\mathrm{I}_t^{\mathrm{New}}$ who become aware of the epidemic threat at day t is determined by a dynamical process where the rate for the individual transi-

tion S → I into the aware state is defined as $\lambda_t^{SN} = \lambda_t^S + \lambda_t^N$, where the terms λ_t^S and λ_t^N represent the social and the news contributions to the force of contagion, respectively. The force of contagion is itself depending on the number of aware individuals I_t and the volume of news M_t at time t. This social-news contagion (SN) model is compartmental (i.e. individuals with the same characteristics are represented by a unique dynamic variable), and takes into account also individuals not yet aware of the disease, and individuals aware but not actively interested in seeking or spreading information on the disease. Mechanisms of waning and re-emergence of the awareness are also considered in the model. For each epidemic, the model's parameters are calibrated separately by using a Markov Chain Monte Carlo (MCMC) approach [64]. The explicit mathematical definition of the model, and the MCMC calibration are detailed in the Material and Methods section of this chapter.

In Fig. 2, 3, 4, we report the results obtained by using the calibrated model to fit the Wikipedia time series. Remarkably, the model is able to capture the large fluctuations induced by the media cycle and the quick rise and decay of attention more typical of the social contagion processes [25]. In order to evaluate the performance of the SN-model, we have considered three other modeling approaches. The first one is a classic regression based on a supervised machine learning approach (L) that models the Wikipedia signal by considering as explanatory variables the news released during the preceding days (see details in the Material and Methods section). We then considered two alternative contagion models, the S-model and the N-model, containing only the and force of contagion, respectively. A closer inspection of these models (reported in the Appendix) shows that each one of these alternative models has clear limitations. For instance, the N-model appears to not reproduce accurately the decay of interest because of saturation in the awareness process [143]. On the contrary, the S-model accounts for the quick rise of awareness but does not captures fluctuations that can be traced back to the media cycle.

In order to put on rigorous ground the comparison across models, we report in Table 1 the basic metrics describing the goodness of fit of the various models (additional metrics in the Appendix section). We observe that the SN-model outperforms all other models on the basic quantities such as the mean absolute percentage error (MAPE), the Pearson correlation coefficient, and the coefficient of determination (R^2).

In Fig. 5 we also report the relative error of each model as a function of time, showing that along the entire time window the SN-model is consistently better performing than the other models. In addition, since we are comparing models with different numbers of parameters, we performed an information theoretic multi-model analysis that clearly shows a very low likelihood that any models could better explain

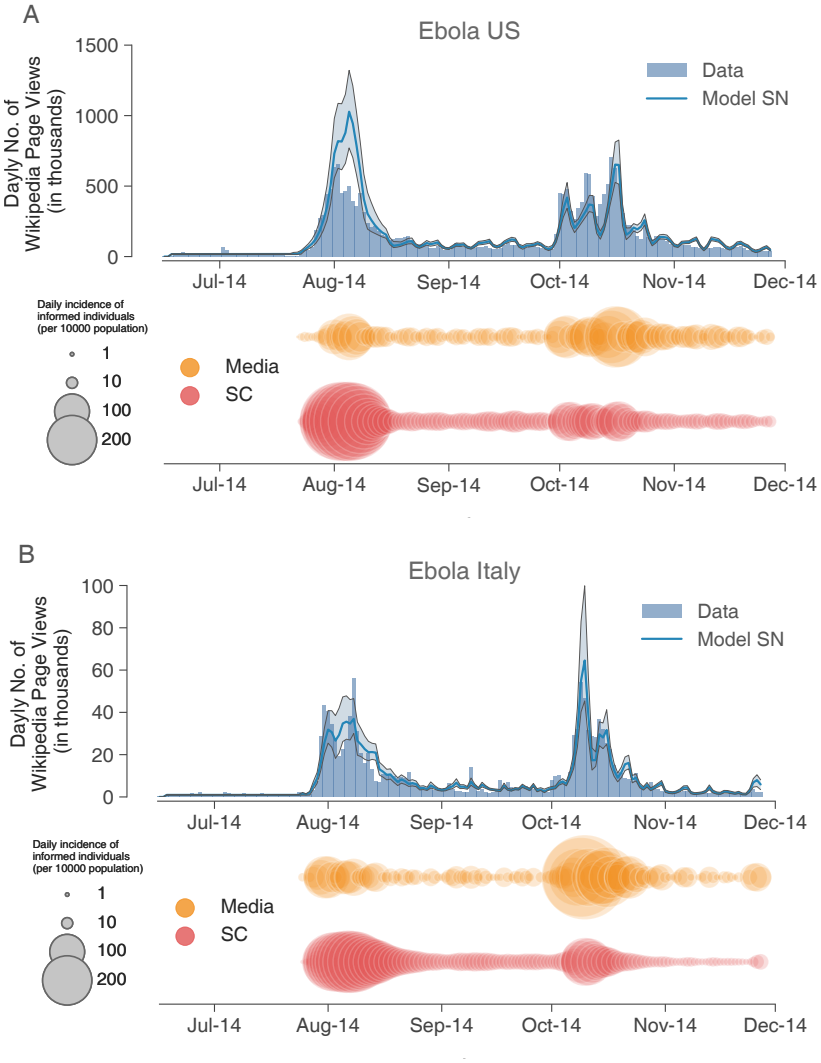

Figure 2: SN-model estimates for Ebola epidemic. **A)** in the United States. **B)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the *SN*-model on the daily number of informed individuals seeking information on Wikipedia. Bubble plots represent the median incidences of informed individuals. Yellow and red bubbles refer to incidences of informed individuals by media communications and through social contagion, respectively.

the data with respect to the *SN*-model [157, 24, 144]. Finally, to assess the predictive power of different modeling approaches here considered, model performances of the *SN*-model were compared with those obtained with the *L*-model, by using only the first 80% of data points (train set) for model calibration and by testing model compliance with the remaining 20% of data (test set). The goodness of fit associated with the two models was compared in terms of MAPE, R^2, and Pearson correlation coefficient, and also compared with the goodness of fit in the last 20% of data points obtained

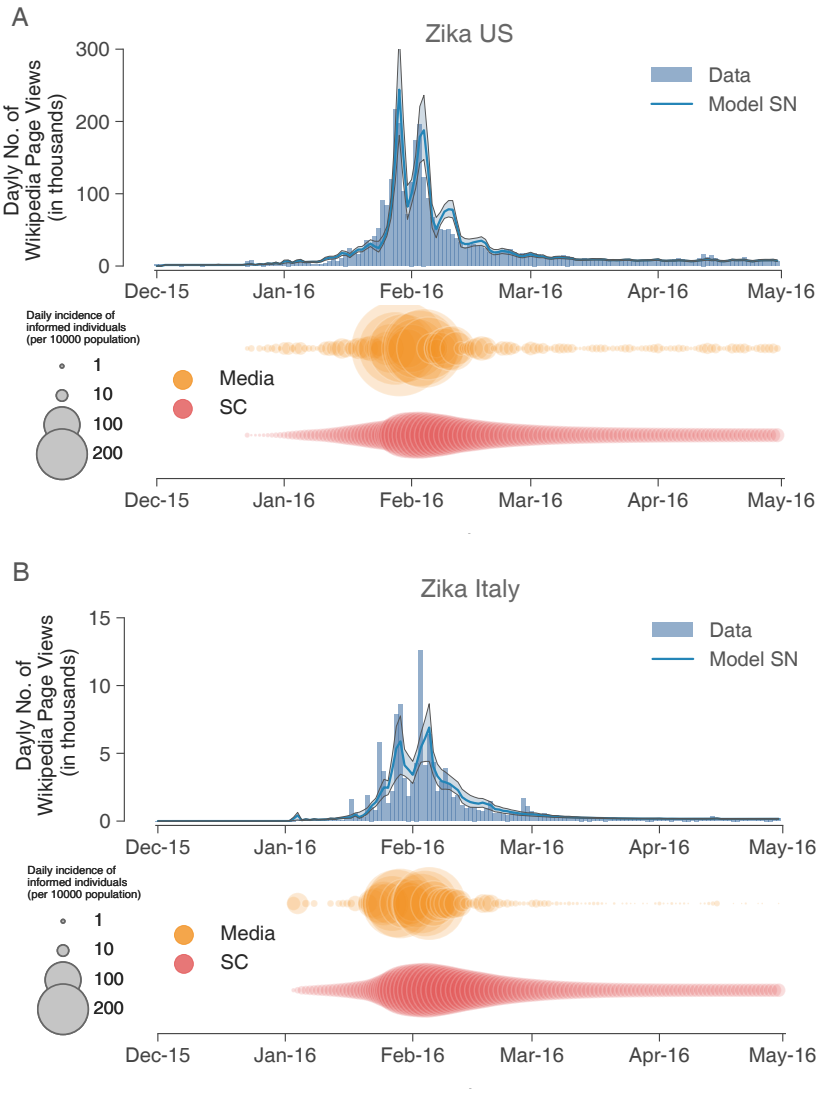

Figure 3: SN-model estimates for Zika epidemic. **A)** in the United States. **B)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the SN-model on the daily number of informed individuals seeking information on Wikipedia. Bubble plots represent the median incidences of informed individuals. Yellow and red bubbles refer to incidences of informed individuals by media communications and through social contagion, respectively.

when using 100% of data for model calibration. The carried out analysis suggests that in all these cases SN-model performs better than the supervised machine learning approach (L, see Table 2).

Our analysis shows that in order to model the spreading of awareness in a population both influence of media and social contagion are relevant mechanisms. A model

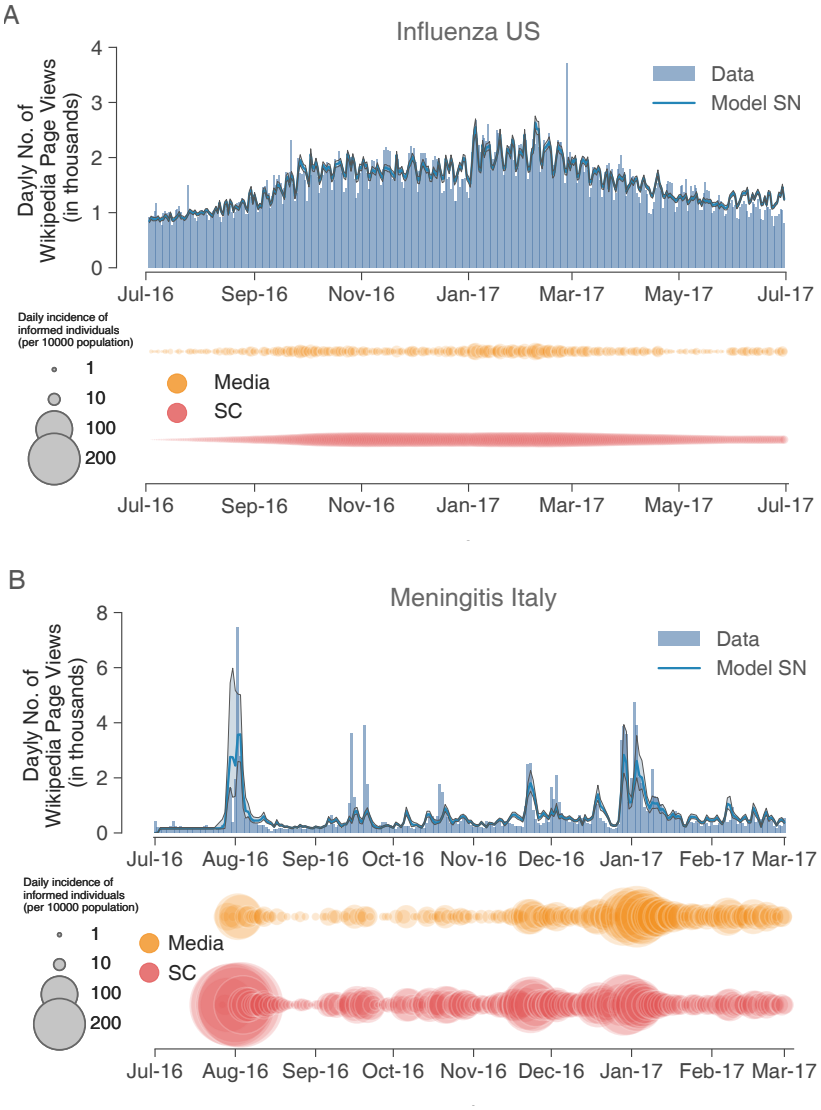

Figure 4: SN-model estimates. **A)** for Influenza in the United States. **B)**for Meningitis in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the *SN*-model on the daily number of informed individuals seeking information on Wikipedia. Bubble plots represent the median incidences of informed individuals. Yellow and red bubbles refer to incidences of informed individuals by media communications and through social contagion, respectively.

accounting for both these routes to develop awareness proves to better reproduce and forecast the dynamics of Wikipedia accesses over time. The model has the added benefit of being mechanistic in describing the impact of media communications and peer-to-peer contagion processes. This allows us to quantify the specific contribution to the process from the two drivers considered and to measure some key features

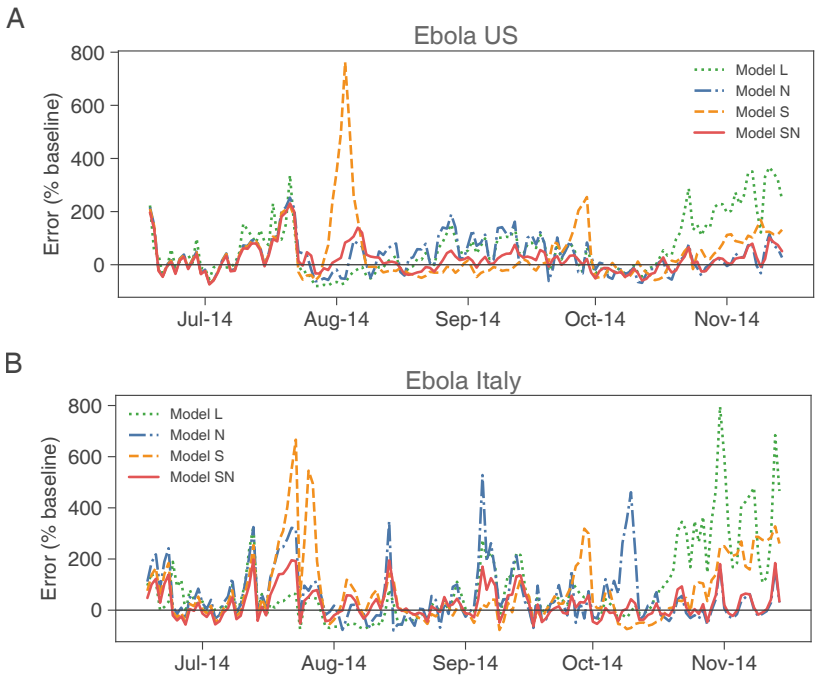

Figure 5: Percentage error between model estimates and data records on the number of Wikipedia page views over time, as obtained for different models when considering the Ebola awareness epidemic. **A)**in the US. **B)** in Italy.

characterizing the spread of awareness in a population.

Our estimates show that, among the different epidemiological scenarios considered, the fraction of individuals that acquired awareness from media broadcast ranges from 30% to 60% (Fig. 6). This result strongly suggests that both media and the word of mouth represent crucial components of the awareness dynamics. Our results are compliant with recent estimates showing that 20% of visits on Wikipedia are triggered by conversations with other individuals and 30% by the media coverage [139].

The estimated proportion of aware people who seek information on Wikipedia (Fig. 6) is expected to mirror the level of interest and concern raised during an epidemic. This proportion was found higher for epidemics declared Public Health Emergencies of International Concerns by the World Health Organization, i.e. Ebola in 2014 (4.6% US, 1.4% Italy) and Zika virus in 2015 (0.9% US, 0.2% Italy), with respect to well known infections as Meningitis (0.08% in Italy between 2016 and 2017) and Influenza (0.14% in US between 2016 and 2017). These estimates suggest that a higher concern is triggered by the emergence of pathogens representing relatively new threats, associated with higher mortality rates [146, 143]. Although the occurrence of severe cases in higher proximity may increase the public attention to a specific disease, the estimated higher impact of Ebola and Zika in US with respect to Italy may be driven by the use of Wikipedia pages written in English by other countries (only 41% of ac-

cess to Wikipedia pages in English are from US; 91% of access to Italian pages are from Italy [53]).

Model estimates of the media transmission rate during different epidemic scenarios (Fig. 6) provide insights on the impact of media communications on different topics across countries. This quantity represents the average number of individuals who get informed by the release of one single news within 24 hours, in a completely uninformed population. Average estimates of this quantity suggest that the impact of media in US might have been more intense (or effective) during the epidemics of Zika in 2015 and Ebola in 2014 with respect to what has occurred during the 2016-17 Influenza season (see Fig. 6) [55, 146, 143]. In Italy a higher media transmission rate was estimated for Ebola in 2014 and Meningitis in 2016-17 with respect to what observed during the 2015 Zika epidemic. Interestingly, for all the scenarios considered but for Influenza in US, the accumulation of media communications over time resulted as an amplifier of the impact of news released afterwards (see Appendix). The negligible effect of past media coverage on the effectiveness of new media communications estimated for Influenza between 2016 and 2017 may be explained by the perceived lower severity of Influenza infection with respect to others and the occurrence of annual regular Influenza epidemics [55].

By comparing estimates of the doubling time characterizing the social contagion mechanism during different epidemics (Fig. 6), i.e. how fast the number of individuals informed though the word of mouth doubles, we found that the spread of awareness through peer-to-peer individual interactions was more than three times faster during 2014 Ebola in the US and 2016-2017 Meningitis in Italy. These two events possibly represent the two threats that have mostly changed the disease risk perception in the public [55, 146, 147].

Finally, as for the Ebola awareness dynamics, the estimated average duration of immunity against the exposure to new information, driving possible waning of individuals' awareness and representing the average time between two visits on a Wikipedia page by the same individual, resulted around 80 days for US and 110 days for Italy. Obtained estimates roughly correspond to the period of time between the two major peaks in the Wikipedia and Google News signals and suggest that two dominant sets of Ebola stories impacted the population in the two countries during 2014. On the opposite, for Zika, and Influenza in US, the estimated duration of immunity exceeds the time frames considered in our analysis, suggesting that the awareness dynamics associated with these two epidemics has been influenced by news stories that has occurred within a relatively narrow period of time. For Meningitis in Italy the average duration of immunity was found around 50 days.

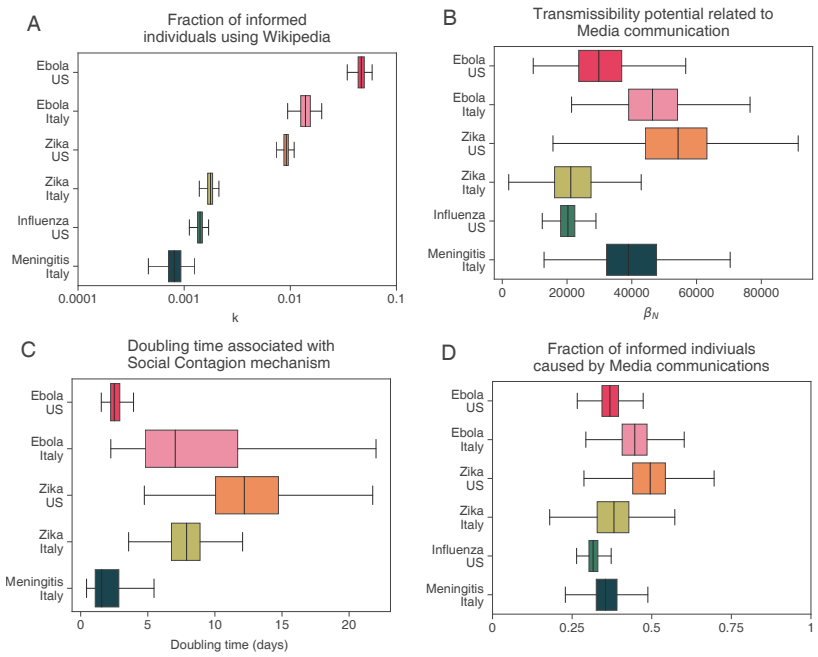

Figure 6: Posterior distribution (2.5%, 25%, 75%, and 97.5% quantiles and mean) of fraction of informed individuals using Wikipedia **(A)**, the transmissibility potential related to media communications **(B)**, the doubling time associated with the social contagion mechanism **(C)**, the fraction of informed individuals due to media communications **(D)**, as obtained for the different epidemic scenarios considered with the SN-model.

2.3 CONCLUSION

The approach presented here provides a modeling framework for investigating time-series related to the spread of awareness of health threats in a population. Our proposed model goes beyond the usual statistical analysis with respect to dependent exogenous variables; indeed, we introduce a mechanistic modeling approach based on the idea that along the news and media drivers, peer-to-peer social contagion plays a major role in the emergence of awareness.

It is worth remarking a number of limitations for the presented approach. First, the models considered for social contagion driven by the word of mouth assume a homogeneous mixing in the population. However, the influence of individuals may be highly heterogeneous and significant different contributions of opinion leaders with respect to the ones of less active individuals may affect the spread of awareness in a population [86, 162]. Clustering of opinions around a given topic within specific population groups or geographical areas are also likely to occur in response to local events or as a consequence of different cultural backgrounds, social norms, and social reinforcement mechanisms [28]. Second, the influence of media is here estimated by using only news released by news websites. Although these are likely rep-

resentative of the overall media attention, we did not considered the heterogeneous contribution of different mass media (e.g. television, radio, newspapers) in shaping the spread of awareness and the impact of news was considered regardless the content of different communications and the reputation of who provides the information. More refined models may explicitly take into account the heterogeneity in the infectiousness associated with different peers (both for media communications and social interactions) instead of adopting a mean field approximation. All these aspects may affect the impact of media on the public, the persistence of concern generated by news and may also be relevant to explain the spread of fake news and misperception [121, 55, 146, 89, 147, 158]. Furthermore, media attention is here considered as an exogenous factor of the considered dynamics, taken as given regardless potential feedback responses between the onset of new stories and people need of more information on a given topic. Finally, our analysis does not provide any indication on whether and how the public has changed their behavior in response to the perceived risk of infection [157, 145, 46, 33, 121]. Further efforts are therefore required to better understand similar mechanisms.

This study however represents a first important attempt to qualitatively and quantitatively describe the role played by the media and the word of mouth in influencing individual awareness and risk perception during different epidemic threats. In particular, the proposed model is able to disentangle the contribution of news influence and social contagion in driving Wikipedia page views in the case of six different public health threats. Furthermore, it provides better explanatory and forecasting power than alternative models considering only one of the driving mechanisms. Most importantly, the model allows the measurement of parameters defining the contagion process such as the fraction of aware people, and the relative contributions of the different contagion processes. The possibility of gathering quantitative information on these parameters is a first step in the operationalization of epidemic models that include the spread of awareness to diseases and the ensuing behavioral changes of the population. The proposed framework is fairy general and can be applied in other contexts related to the diffusion of information and knowledge.

2.4 MATERIAL AND METHODS

2.4.1 *Data Description*

The daily numbers of visits to Wikipedia pages on specific diseases were obtained by publicly available data [51, 52] and used to model temporal changes in the number of individuals seeking information on a specific epidemic threat. Pages in Italian and English, accessed between 2014 and 2017, were used for Italian and US users respec-

tively. Wikipedia data were preferred to other public available datasets (e.g. Google Trends Data, etc.) for the following reasons: a) Wikipedia visits may better reflect the need to acquire in-depth information about a diseases, instead of capturing recurrent accesses to an updated source of information on what is occurring during an epidemic; and b) access of individuals are provided in absolute numbers, therefore allowing a comparison of time-series associated with different countries and diseases. Accesses during 2016 on any Wikipedia page in two languages were considered to assess, separately for the two countries, potential differences in the individual use of Wikipedia across different days of the week (e.g. weekdays/weekend). Deviations with respect to weekly means were used to remove spurious fluctuations in Wikipedia accesses (details can be found in the Appendix section). Temporal changes of media attention to a given epidemic were modeled by using Google News platform [68]. Specifically, a proxy of daily media response to potential threats was defined by the amount of articles released by websites based in the country containing the name of the considered epidemic in its headline. A cross-correlation at different time lags was performed to assess potential synchrony in the two types of signals.

2.4.2 *Supervised machine learning approach*

We conduct a multi-linear regression analysis (L) in order to test whether the dynamics of Wikipedia accesses mirror changes of the media attention to a given disease. We consider the number of Wikipedia page views at time t as the response variable and the amounts of news released at different times prior t as potential explanatory variables. Specifically a set of linear models was defined as follows:

$$\mathrm{W}_t = \sum_{i=0}^{T} \alpha_i \mathrm{M}_{t-i} + \mathrm{W}_0,$$

where W_t represents the daily number of Wikipedia page views at time t, α_i are the regression coefficients, M_t is the number of news released at time t, T defines the number of days before t in which media communication can affect Wikipedia accesses and W_0 reflects accesses to Wikipedia in the absence of media coverage on a given disease. Best values for T were obtained as a result of a model selection procedure based on the Akaike information criterion [24]. Details are reported in the Appendix.

2.4.3 *Modeling awareness as a contagion process*

Word of mouth and media communications are both considered as plausible routes of transmission. The spread of information through the word of mouth was modeled by assuming homogeneous-mixing in the population and a force of infection

$$\lambda_t^S = \frac{\beta_S}{\mathrm{N}} \mathrm{I}_t,$$

where β_S represents the transmission rate for the social contagion associated with peer-to-peer conversations between individuals.
As for the rate at which individuals becomes aware thanks to media communication, we assume that the impact of news released at time t can be inflated by news released in the previous days. The strength of media communications to reach an individual is therefore defined as follows:

$$\lambda_t^N = \frac{\beta_N}{\mathrm{N}} \mathrm{M}_t \left(\prod_{i=t_0}^{t-1} \left(1 + \mathrm{M}_i e^{-\rho(t-i)} \right) \right)$$

, where β_N is the transmissibility potential related to media coverage at time t, while the term in parenthesis defines the amplifying mechanism due to past media communications. Specifically, the contribution of past communications to the amplification of the strength of news released afterwards is assumed to exponentially decay with time passed since their release: smaller the value of ρ, longer the influence of past communications. For large value of ρ , the model considers only media communication released at time t, i.e. $\lambda_t^N = \frac{\beta_N}{\mathrm{N}} \mathrm{M}_t$. Alternative amplification mechanisms are considered in the SI. The two routes of transmission are considered alone (models S and N) or combined in a nested model (SN). Briefly, in the proposed models, transitions between classes can be described by the following system of ordinary differential equations: $d\mathrm{S}/dt = -\lambda(t)\mathrm{S}(t) + \nu\mathrm{R}(t)$; $d\mathrm{I}/dt = +\lambda(t)\mathrm{S}(t) - \gamma\mathrm{I}(t)$; $d\mathrm{R}/dt = +\gamma(t)\mathrm{I}(t) - \nu\mathrm{R}(t)$, where $\lambda(t)$ represents the force of contagion at time t; $1/\gamma$ is the average time period in which an aware individual can spread the information through word of mouth; $1/\nu$ is the average duration of immunity against the exposure to new information. Such a period of time can be interpreted as the average time between two visits on a Wikipedia page by the same individual. The proposed SIRS model, for specific choices of free model parameters, can degenerate into a SIR model (no immunity waning), and even into a SI model (no immunity).

2.4.4 *Model calibration and goodness of fit*

Each model was calibrated separately. Free model parameters were estimated by using a Markov Chain Monte Carlo approach [64] applied to the negative binomial Likelihood of the observed Wikipedia page views for each scenario. Goodness of fit was assessed through a wide set of statistical measures including, among others, the Akaike and Deviance Information Criteria (AIC, DIC), the mean absolute percentage error (MAPE) and the Pearson correlation coefficient [157, 24, 144]. Details can be found in the Appendix section.

		MAPE	Pearson	R^2	AIC	P. of I.L.
Ebola US	*L*	84.895	0.666***	0.352	3709.75	< 0.001
	N	56.883	0.832***	0.659	3635.1	< 0.001
	S	70.214	0.584***	-8.853	3648.48	< 0.001
	SN	39.597	0.844***	0.551	3556.51	-
Ebola IT	*L*	109.63	0.61***	0.289	2831.26	< 0.001
	N	80.224	0.706***	-4.744	2792.92	< 0.001
	S	97.871	0.729***	0.509	2808.03	< 0.001
	SN	46.521	0.913***	0.809	2681.16	-
Zika US	*L*	72.545	0.787***	0.523	3259.58	< 0.001
	N	52.509	0.932***	0.846	3173.02	< 0.001
	S	33.393	0.89***	0.775	3080.87	< 0.001
	SN	30.482	0.929***	0.841	3059.46	-
Zika IT	*L*	203.127	0.782***	0.492	2017.67	< 0.001
	N	374.899	0.874***	0.71	2119.44	< 0.001
	S	58.985	0.805***	0.041	1911.98	< 0.001
	SN	56.774	0.871***	0.746	1853.94	-
Meningitis IT	*L*	71.238	0.569***	0.311	3392.09	< 0.001
	N	57.034	0.792***	0.588	3333.81	< 0.001
	S	74.089	0.524***	0.161	3369.43	< 0.001
	SN	51.203	0.739***	0.526	3280.18	-
Influenza US	*L*	14.066	0.781***	0.589	4907.52	< 0.001
	N	16.772	0.719***	0.503	4997.7	< 0.001
	S	17.231	0.706***	0.49	5000.99	< 0.001
	SN	10.138	0.88***	0.773	4693.88	-

*****p value < 0.001**

Table 1: Basic metrics describing the goodness of fit associated with different epidemics and models including the mean absolute percentage error (MAPE), the Pearson correlation coefficient, the coefficient of determination (R^2) and the Akaike information criterion (AIC). Values of R^2 were computed on the basis of equation y=ax, thus allowing for negative R^2 values. AIC was used to estimate the probability that information loss (P. of I.L.) is minimized when we consider an alternative model to the model having the lowest AIC value.

		MAPE		Pearson		R^2	
		80/20	**Baseline**	**80/20**	**Baseline**	**80/20**	**Baseline**
Ebola US	L	258.46	188.22	0.85***	0.85***	-2.08	-0.53
	SN	46.30	31.93	0.91***	0.92***	0.70	0.83
Ebola IT	L	382.95	291.96	0.77***	0.78***	-3.74	-1.58
	SN	46.91	38.88	0.94***	0.94***	0.81	0.86
Zika US	L	122.38	35.01	0.50*	0.53*	-17.06	-0.72
	SN	53.61	30.23	0.90***	0.87***	-4.59	-1.00
Zika IT	L	121.08	47.77	0.57**	0.63**	-1.59	0.34
	SN	42.38	40.64	0.90***	0.90***	0.27	0.64
Meningitis IT	L	230.92	124.91	0.77***	0.77***	-10.57	-1.99
	SN	69.77	41.53	0.49**	0.67***	-0.27	0.40
Influenza US	L	21.48	19.91	0.26**	0.27**	-1.56	-1.22
	SN	16.92	13.75	0.48***	0.48***	-0.29	0.09

***p value < 0.1 **p value < 0.05 ***p value < 0.001**

Table 2: Statistical measures on the performances of the SN-Model (SN) and the supervised machine learning approach (L), as obtained for different epidemic scenarios. Each measure was obtained for two distinct calibration procedures. In the first one (labeled as 80/20 in the table), model parameters were calibrated using only 80% of data; in the second (baseline) model parameters were calibrated using 100% of data. In both cases, performances were assessed only in the last 20% data points.

APPENDIX

2.A SUPPORTING MATERIAL

2.A.1 *Data description*

Four different epidemic threats were considered in our study:

1. the Ebola virus epidemic in West Africa during 2014;
2. the Zika virus epidemic between 2015 and 2016;
3. the Meningitis outbreak in Italy between 2016 and 2017;
4. the seasonal Influenza in US between 2016 and 2017.

In our analyses, Wikipedia page views associated with different diseases were used as a proxy of awareness in the population during the corresponding epidemic. The number of news released in the Internet was used as a proxy of media attention in the same period of time.
Wikipedia page views for Meningitis and Influenza were obtained by using a free tool, recently made available through the community-led Wikimedia Toolforge project, for the analysis of the number of page views of published articles [52]. Specifically, daily desktop views on Influenza page in English and on Meningitis page in Italian were considered.
As for Ebola and Zika, new Wikipedia pages were created as a consequence of the emergence of a new epidemic threat. Hence, we considered multiple pages for these two epidemics. More precisely, each page including the terms “Ebola” and “Zika” in the page title, and the respective redirects incoming to that pages were considered. Data for page views in English and Italian were retrieved from a dataset providing hourly aggregates of page views to any desktop version of Wikipedia pages [51] and aggregated at the day level afterwards.

We tested whether spurious temporal fluctuations in Wikipedia page views may have occurred due to a different use of Wikipedia across different days of the week. To this aim, for each week in 2016, we computed the number of accesses to all pages of Wikipedia associated with different days of the week and the corresponding weekly mean. The empirical distribution of deviations from the weekly mean for 52 weeks

were defined as the ensemble of deviations $D(d,w)$ associated with a given day $d = \{\text{monday},\ldots,\text{sunday}\}$ of the all weeks $w = \{1,\ldots,52\}$ in 2016:

$$D(d,w) = \frac{\omega(d,w) - \overline{\omega}(w)}{\overline{\omega}(w)},$$

where $\omega(d,w)$ is the number of visits in the day d of week w and $\overline{\omega}(w)$ is the average of daily number of visits during the week w. Deviations were computed separately for Wikipedia in Italian and English and are shown in Fig. 7. The number of page

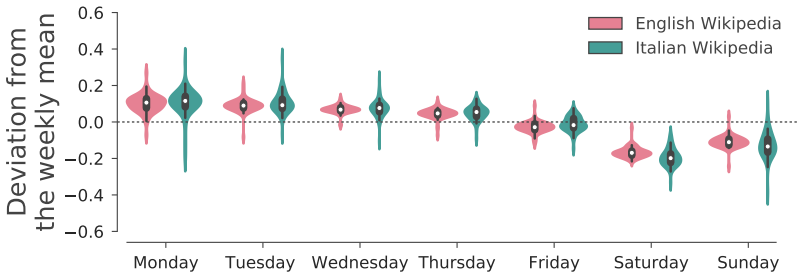

Figure 7: Violin plots of the distribution of the deviations in the number of visits associated with a given day of the week from the average number of visits recorded in the same week, as detected across different weeks of 2016.

views associated with a given day of the week was also compared with the average number of page views in the remaining days of the week. The intrinsic variation in the Wikipedia visits across different days of the week in 2016 was found statistically significant (t-test) for 6 out of 7 days of the week (Table 3). Thus, we have decided to analyze the temporal changes in the awareness of individuals on a given epidemic threat by removing from the original raw signal ω_{raw}, defined as the observed number of page views over time, the intrinsic weekly seasonality identified by the analysis of weekly deviations during 2016. We therefore re-computed the number of Wikipedia page views as

$$\omega(t) = \frac{\omega_{\text{raw}}(t)}{(1+\overline{D}_{d(t)})},$$

where $d(t)$ is the day of the week associated with time t and $\overline{D}_d$ is the estimated mean of the distribution of the deviation as obtained by analyzing data on access to all pages of Wikipedia in 2016 [52]. A comparison between $\omega(t)$ and $\omega_{\text{raw}}(t)$ is shown in Fig. 8.

Figure 9 shows the two datasets used in our analysis.

Temporal changes in the media coverage on a given topic were estimated on the basis of the daily number of news released, as obtained by using Google News platform [68]. Google News is a news aggregation service provided free-of-charge, that collects thousands of articles published in news websites, providing information on the date and country in which news are released. Articles containing the name of the

	English		Italian	
	t-test	**p-value**	**t-test**	**p-value**
Monday vs. others	6.89	<0.001	5.85	<0.001
Tuesday vs. others	6.18	<0.001	5.91	<0.001
Wednesday vs. others	4.70	<0.001	4.27	<0.001
Thursday vs. others	3.09	0.002	2.80	0.005
Friday vs. others	-1.94	0.052	-0.70	0.48
Saturday vs. others	-11.50	<0.001	-11.12	<0.001
Sunday vs. others	-7.50	<0.001	-7.03	<0.001

Table 3: Student's t-test results as obtained by comparing, separately for each day of the week (Monday,..., Sunday), the average number of page views occurred in a given day of the week for any week in 2016 with the average number of page views during the remaining days of the same week.

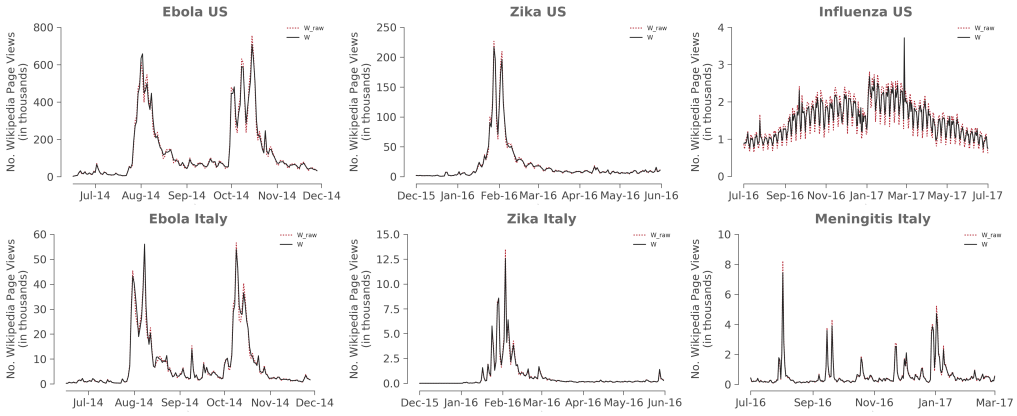

Figure 8: Comparison between the original raw signal ω_{raw} and the adjusted signal ω for the six epidemic scenarios under analysis. ω was obtained by removing the expected oscillations in numbers of page views due to intrinsic variations across different days of the week, as identified for all Wikipedia pages' visits during 2016

infection in their headlines were considered. Specifically, "Ebola", "Zika" and "Meningite" (denoting Meningitis in Italian) were used as search keyword for the corresponding epidemic threats. News on Ebola and Zika were retrieved separately for news released in Italy and in US. Finally, both terms "Influenza" and "flu" were included to

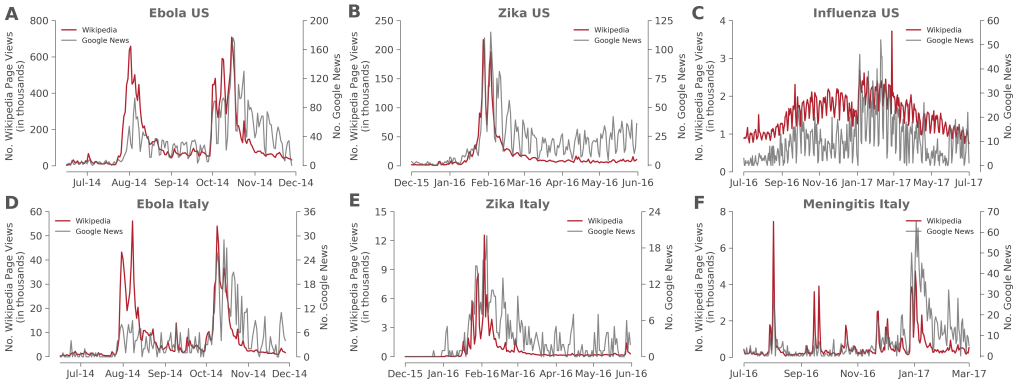

Figure 9: Data (adjusted signals) on the amount of Wikipedia page views and news from Google News platform for Ebola **(A)**, Zika **(B)**, Influenza **(C)** in the US and for Ebola **(D)**, Zika **(E)**, Meningitis **(F)** in Italy.

select articles related to Influenza published in US.

Possible time lags between the two signals (Google News vs. Wikipedia) were evaluated by performing a Cross-correlation analysis. Obtained results suggest an overall synchrony between the media coverage and the individuals' awareness about a specific disease threat. The time series associated with these two data sources were found significantly correlated (Pearson p-value <0.001) for all the considered scenarios (Figure 10 and Table 4).

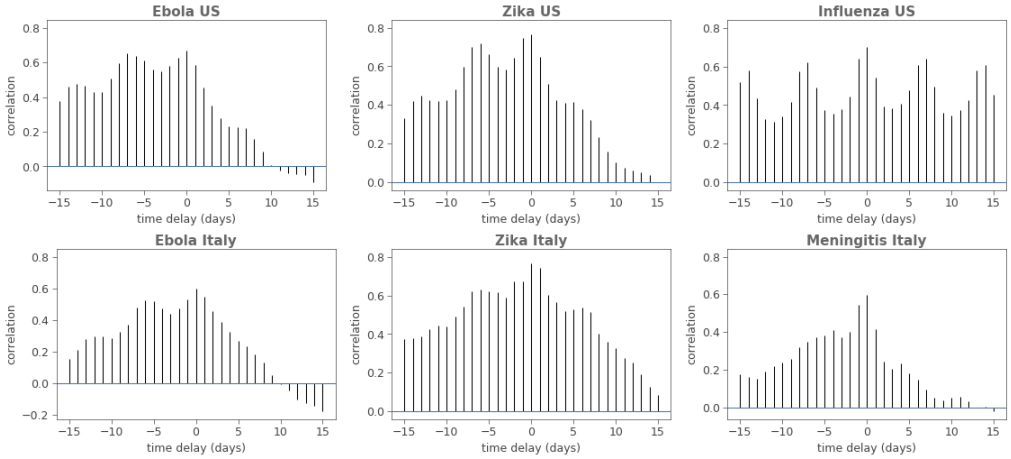

Figure 10: Cross correlation at different time lags (in days) of time series of Wikipedia page views and the number of news collected by Google News platform for the six epidemic scenarios considered.

	Pearson correlation coefficient	**Spearman correlation coefficient**
Ebola US	0.67***	0.79***
Ebola Italy	0.60***	0.72***
Zika US	0.77***	0.73***
Zika Italy	0.76***	0.82***
Influenza US	0.60***	0.65***
Meningitis Italy	0.70***	0.72***

*****p value < 0.001**

Table 4: Pearson and Spearman correlations between the daily number of news collected by the Google News platform and the number of Wikipedia page views for the six epidemic scenarios considered.

2.A.2 *Supervised machine learning approach*

A regression analysis was conducted by using Wikipedia visits at time t as target variables, and the count of Google News items released before t as explanatory variables. Specifically, in this analysis we assume that the number of Wikipedia visits over time can be approximated as follows:

$$\mathrm{W}_t = \sum_{i=0}^{T} \alpha_i \mathrm{M}_{t-i} + \mathrm{W}_0$$

where W_t is the estimated number of Wikipedia page views at time t; M_t is the amount of news released at time t; α_i is the regression coefficient associated with the explanatory variable M_{t-i}; W_0 represents the expected number of Wikipedia page views in absence of media attention; T defines the number of days before t during which media communications influence the number of Wikipedia visits at time t. In particular, when we consider $T = 0$ we are assuming that the number of Wikipedia visits at time t only depends on the number of news released at time t. Values of T considered in our analysis range from 0 to 9, so that ten different multi-linear models were applied to each epidemic scenario and calibrated separately.

For each epidemic scenario, the best regression model was defined on the basis of the Akaike information criterion (AIC). More specifically, AIC for a given model m was computed as the average of all AIC values associated with 30,000 simulations obtained by sampling parameters from their posterior distribution as obtained by a Monte Carlo Markov Chain (MCMC) approach:

$$\mathrm{AIC}(m) = \frac{1}{J} \sum_{i=0}^{J} 2K(m) - 2\ln(L_j),$$

where K represents the number of parameters of model m; J is the number of samples drawn from the posterior distribution of free model parameters; L_j is the Likelihood associated with the j^{th} sample. According to the adopted criterion, best models are defined as those associated with lower values of AIC. Once the best model b is selected, AIC can be used to estimate the probability p that the information loss is minimized when we adopt an alternative model i instead of model b [24]:

$$p = \exp\left(\text{AIC}_b - \text{AIC}_i/2\right),$$

where AIC_b and AIC_i are the AIC values associated with the best model b and the alternative candidate model i, respectively.
Obtained results are shown in Tables 5 and 6. A trade-off between the number of parameters and the goodness of fit emerges in the AIC values obtained with different models considered. This result is due to the fact that the AIC measure includes a penalty term for increasing the number of free model parameters ($2K$).

Threshold time	**Number of parameters**	**AIC**					
T	K	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
0	3	3719.36	2891.59	3272.69	2129.98	3398.5	4994.24
1	4	3709.99	2841.37	3261.05	2104.29	3392.09*	4986.13
2	5	3709.75*	2831.96	3259.58*	2089.49	3396.67	4974.82
3	6	3712.07	2831.26*	3261.53	2017.67*	3401.25	4965.7
4	7	3715.48	2833.74	3264.66	2021.58	3404.8	4958.87
5	8	3719.73	2836.16	3267.37	2024.76	3409.43	4936.12
6	9	3724.84	2840.46	3270.81	2024.76	3414.63	4914.43
7	10	3729.3	2844.7	3274.91	2027.17	3417.62	4907.52*
8	11	3733.42	2848.72	3278.35	2030.46	3423.12	4908.57
9	12	3738.05	2852.49	3281.95	2030.01	3427.49	4910.34

* model associated with best performances according to the considered statistical measure

Table 5: AIC values obtained with the different models considered.

2.A.3 *Contagion models*

All contagion models considered in our analysis to describe the spread of awareness in the population are based on a SIRS transmission schema [6]. In these models, the population is divided into three different compartments: individuals that are not aware about the epidemic and can get informed either by social interactions with

Threshold time	**Number of parameters**	**Probability that information loss is minimized**					
T	K	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
0	3	0.008	< 0.001	0.001	<0.001	0.041	< 0.001
1	4	0.89	0.006	0.48	<0.001	1*	< 0.001
2	5	1*	0.702	1*	<0.001	0.101	< 0.001
3	6	0.314	1*	0.377	1*	0.01	< 0.001
4	7	0.057	0.289	0.079	0.141	0.002	< 0.001
5	8	0.007	0.086	0.02	0.029	< 0.001	< 0.001
6	9	0.001	0.01	0.004	0.029	< 0.001	0.032
7	10	< 0.001	0.001	< 0.001	0.009	< 0.001	1*
8	11	< 0.001	< 0.001	< 0.001	0.002	< 0.001	0.592
9	12	< 0.001	< 0.001	< 0.001	0.002	< 0.001	0.244

* model associated with best performances according to the considered statistical measure

Table 6: Probability that the information loss is minimized when we adopt an alternative model instead of the one associated with the lowest AIC value [24].

other individuals or by media communications (S); individuals who are aware about the ongoing outbreak and can share their concern and information with other individuals (I), and people who were previously informed about the epidemic and are immune against the exposure to new information (R). Transitions between classes can be described by the following system of ordinary differential equations:

$$\begin{cases} \dfrac{d\mathrm{S}(t)}{dt} = -\lambda(t)\mathrm{S}(t) + \nu\mathrm{R}(t) \\ \dfrac{d\mathrm{I}(t)}{dt} = +\lambda(t)\mathrm{S}(t) - \gamma\mathrm{I}(t) \\ \dfrac{d\mathrm{R}(t)}{dt} = \gamma\mathrm{I}(t) - \nu\mathrm{R}(t) \end{cases} \tag{1}$$

where t denotes time, $\lambda(t)$ represents the rate at which individuals get informed at time t by conversations with other individuals or media communications; $1/\gamma$ is the average time period in which an aware individual can contribute to spread the information about the epidemic threat through word of mouth and represent the average time required to develop immunity against the exposure to new information; finally, $1/\nu$ represents the average duration of such immunity. In our model formulation, variables S, I, R reflect respectively the fraction of susceptible, infected and

immune individuals in the population in a classic SIRS epidemic schema. Similarly, $1/\gamma$, $1/\nu$ and $\lambda(t)$ mirror the average duration of infectiousness, the average duration of immunity against re-infection and the force of infection over time [6].
In particular, we define λ_t as the sum of two distinct contributions:
i) λ_t^S representing the contribution to the overall force of infection due to social contagion and depending on the number of aware individuals at time t, i.e. $\lambda_t^S = \lambda^S(\mathrm{I}(t))$;
ii) λ_t^N representing the contribution to the overall force of infection due to media response to the epidemic and modeled as a function of media communications occurred before time t, i.e. $\lambda_t^N = \lambda^N(\mathrm{M}(t))$.

We assume that the spread of information through social contagion is driven by homogeneous mixing within the considered population. In this framework, homogeneous mixing approach indicates that each individual can get informed by any aware individual in the population, This means that possible social structures and spatial heterogeneities affecting the word of mouth are not considered in our models.
Specifically, we defined $\lambda^S(\mathrm{I}_t)$ as follows:

$$\lambda^S(\mathrm{I}_t) = \frac{\beta_S}{\mathrm{N}}\mathrm{I}_t,$$

where N represents the number of individuals in the considered population and β_S defines the transmission rate associated with peer-to-peer conversations.

As for the awareness generated by the media response to the epidemic threat, different formulations of $\lambda^N(\mathrm{M}_t)$ were considered. The first one assumes that only news released at time t, namely M_t, can change the status of individual awareness at time t.
In this case is defined as:

$$\lambda_t^N = \frac{\beta_N}{\mathrm{N}}\mathrm{M}_t, \tag{2}$$

where β_N is the transmission rate associated with one single media communication released at time t; M_t is the number of news released at time t.

Other two models were considered for $\lambda^N(\mathrm{M}_t)$, taking into account the potential effect of past communications on individual response to new communications. The first model we considered for $\lambda^N(\mathrm{M}_t)$ is defined as follows:

$$\lambda_t^N = \frac{\beta_N}{\mathrm{N}}\sum_{i=t_0}^{t}\mathrm{M}_i e^{-\rho(t-i)}, \tag{3}$$

where ρ is a parameter driving the waning of individuals' memory about past communications; t_0 is the time at which the awareness of a given epidemic starts to spread in the population; M_i is the number of news released at time i; β_N represents the

transmission rate associated with one single media communication released at time t when $M_i = 0$ for any i before t, i.e. when $\lambda_t^N = \frac{\beta_N}{N} M_t$. It is worth noting that, in principle, if ρ is sufficiently large this formulation of $\lambda^N(M_t)$ reduces to $\lambda_t^N = \frac{\beta_N}{N} M_t$.

The second model considered for $\lambda^N(M_t)$ is defined as follows:

$$\lambda_t^N = \frac{\beta_N}{N} M_t \left(\prod_{i=t_0}^{t-1} 1 + M_i e^{-\rho(t-i)} \right), \tag{4}$$

where, again, ρ is a parameter driving the waning of individuals' memory about past communications; t_0 is the time at which the awareness of a given epidemic starts to spread in the population; M_i is the number of news released at time i; β_N represents the transmission rate associated with one single media communication released at time t when $M_i = 0$ for any i before t, i.e. when $\lambda_t^N = \frac{\beta_N}{N} M_t$. For ρ is sufficiently large this formulation also reduces to $\lambda_t^N = \frac{\beta_N}{N} M_t$.

However, in model 3, the contribution of media communications released at time t is not affected by news released before t, which in turn independently contribute to the overall force of infection at time t. On the opposite, in model 4, the contribution of communications released at time t is directly amplified by the amount of news released before t. Contagion models considered in our analysis are summarized in Table 7, which includes for each model the list of corresponding free model parameters.

When we consider that the awareness of individuals spread through a contagion process, we assume that the number of Wikipedia page views at time t associated with a specific disease, namely W_t, reflects the number of individuals who get informed about the epidemic threat at day t, namely I_t^{New}. Specifically, we assume that the number of Wikipedia page views over time can be approximated as:

$$W_t = k I_t^{New} + W_0,$$

where $k \in [0,1]$ represents the fraction of informed individuals seeking information in Wikipedia and W_0 represents the expected number of Wikipedia page views in absence of any information on the ongoing epidemic threat. Initial conditions for system 1 were defined as $S_{t=t_0} = N, I_{t=t_0} = R_{t=t_0} = 0$, where t_0 represents the time at which the awareness of a given epidemic starts to spread in the population. Finally, we assume that the number of individuals who get informed by an epidemic at t_0 is $I_0^{New} = \frac{\omega_0 - W_0}{k}$. In our analysis the stochastic version of models described by system 1 were simulated.

2.A.4 *Models calibration*

The estimation of free model parameters was carried out, separately for each scenario and each candidate model, through a MCMC approach with random walk Metropolis-

		Force of contagion	
Model	**Free parameters**	**Social contagion** λ_t^S	**Media response** λ_t^N
S	$k, \beta_S, \gamma, \nu, t_0, \mathrm{W}_0, x$	$\frac{\beta_S}{\mathrm{N}}\mathrm{I}_t$	0
N$^{\#}$	$k, \beta_N, \nu, t_0, \mathrm{W}_0, x$	0	$\frac{\beta_N}{\mathrm{N}}\mathrm{M}_t$
N$^{\#\#}$	$k, \beta_N, \rho, \nu, t_0, \mathrm{W}_0, x$	0	$\frac{\beta_N}{\mathrm{N}}\sum_{i=t_0}^{t}\mathrm{M}_i e^{-\rho(t-i)}$
N$^{\#\#\#}$	$k, \beta_N, \rho, \nu, t_0, \mathrm{W}_0, x$	0	$\frac{\beta_N}{\mathrm{N}}\mathrm{M}_t\left(\prod_{i=t_0}^{t-1} 1+\mathrm{M}_i e^{-\rho(t-i)}\right)$
SN$^{\#}$	$k, \beta_N, \beta_S, \gamma, \nu, t_0, \mathrm{W}_0, x$	$\frac{\beta_S}{\mathrm{N}}\mathrm{I}_t$	$\frac{\beta_N}{\mathrm{N}}\mathrm{M}_t$
SN$^{\#\#}$	$k, \beta_N, \beta_S, \gamma, \rho, \nu, t_0, \mathrm{W}_0, x$	$\frac{\beta_S}{\mathrm{N}}\mathrm{I}_t$	$\frac{\beta_N}{\mathrm{N}}\sum_{i=t_0}^{t}\mathrm{M}_i e^{-\rho(t-i)}$
SN*	$k, \beta_N, \beta_S, \gamma, \rho, \nu, t_0, \mathrm{W}_0, x$	$\frac{\beta_S}{\mathrm{N}}\mathrm{I}_t$	$\frac{\beta_N}{\mathrm{N}}\mathrm{M}_t\left(\prod_{i=t_0}^{t-1} 1+\mathrm{M}_i e^{-\rho(t-i)}\right)$

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 7: Summarized description of contagion models considered in our analysis.

Hastings sampling [64], applied to the negative binomial likelihood of observing the daily number of Wikipedia page views related to the considered disease and country. The likelihood associated with a specific parameter set θ was defined as:

$$L(\theta, x|\omega_d) = \prod_d \frac{\Gamma(x+\omega_d)}{\Gamma(x)\omega_d!}\left(\frac{\mathrm{W}_d(\theta)}{\mathrm{W}_d(\theta)+x}\right)^{\omega_d}\left(\frac{x}{\mathrm{W}_d(\theta)+x}\right)^{x}$$

where d runs over the days considered to calibrate models for a specific epidemic threat; ω_d is the number of observed Wikipedia page views occurred within day d; x is the dispersion parameter of the negative binomial distribution; $\Gamma(\cdot)$ is the standard gamma function; $\mathrm{W}_d(\theta)$ is the estimated number of Wikipedia page views at the day d as obtained with the considered model and parameter set θ.

At each iteration, the algorithm evaluates the likelihood of a new candidate vector of parameters, which is accepted or not based on the standard Metropolis-Hastings algorithm. Uniform prior distributions are assumed for the free parameters; new candidate values of free model parameters are proposed from a normal distribution centered on their current values. MCMC convergence was assessed by checking that, after a burn-in period of 2,000 iterations, the trace plots associated with different chains, i.e. the sequence of accepted parameter values, were approximately characterized by a constant width and average, therefore proving good mixing of the parameters. Figure 11 shows the trace plots associated with illustrative chains consisting of 30,000 iterations, as obtained through the MCMC approach with random walk

Metropolis–Hastings sampling applied to the negative binomial likelihood of the observed Wikipedia page views on Ebola in US by using model SN based on equation 4.

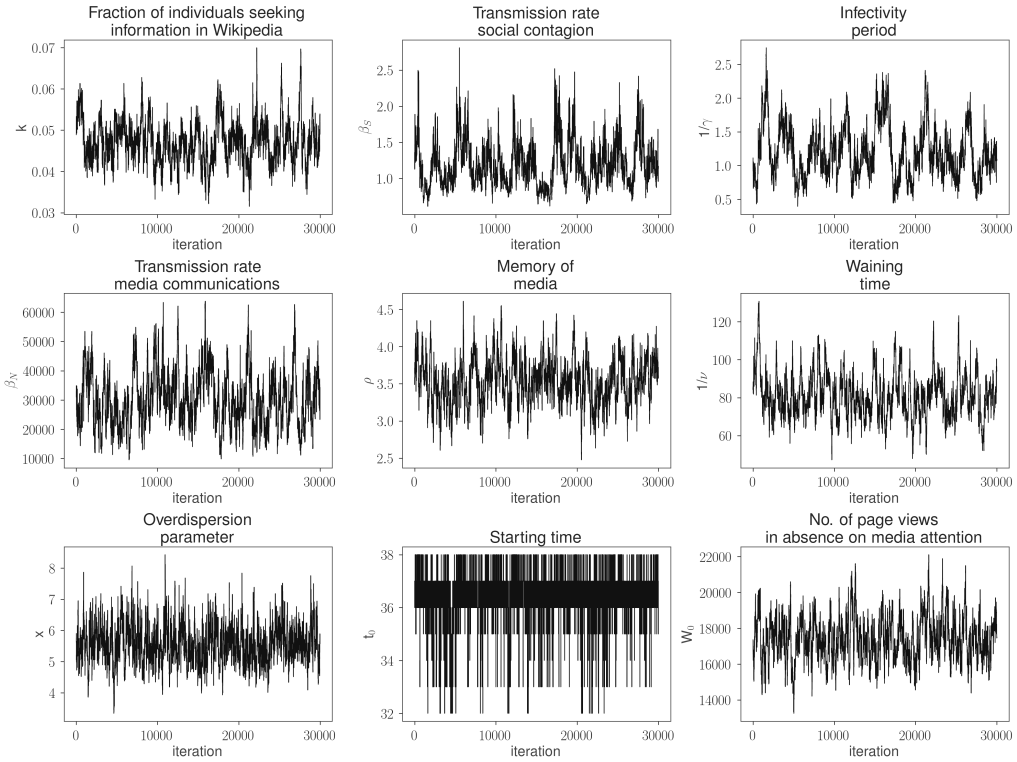

Figure 11: Trace plots of illustrative MCMC chains as obtained with random walk Metropolis–Hastings sampling applied to the negative binomial likelihood of the observed Wikipedia page views on Ebola in US by using model SN based on equation 4

2.B ADDITIONAL RESULTS

2.B.1 *Model performances*

Accuracy of different models in reproducing the observed data was assessed by using a variety of statistical measures [157]. These include the root mean square error between model estimates and data, the corresponding mean absolute error (MAE) and mean absolute percentage error (MAPE), the coefficient of determination (R^2), the Pearson correlation coefficient, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Deviance Information Criterion (DIC). Specifically, for each model and epidemic scenario considered, all the statistical measures, with the exception of the DIC, were computed as the average of the 30,000 values of the considered statistic as obtained by running 30,000 simulations with parameter sets sampled from the posterior distribution of parameters estimated with the MCMC approach. In addition to all contagion models summarized in Table 7, the

best regression model (denoted by L) was included in this analysis. Obtained results associated with different epidemic scenarios and models are shown in Table 8 9 10 11 12 13 14 15 16. The corresponding model estimates on the number of Wikipedia page views over time are shown in Figures 12 13 14 15 16 17 18 19.

According to the majority of the statistical measures here considered, the nested model (SN) outperforms all other models. Parameters estimates obtained with model SN based on equation 4 are reported in Table 17. Although available data do not supply sufficient information for reliably estimating all free parameter values (e.g., for estimating both β_S and $1/\gamma$), interesting insights emerge from the comparison of estimates obtained across different scenarios considered (e.g., when comparing the doubling time associated with the social contagion during different epidemic threats).

Finally, model performances were also measured by using only 80% of data points (train-set) for model calibration and by testing the model compliance with the remaining 20% (test-set). Only two models were considered for this analysis:
1) model SN, representing the best model according to different statistical measures of the goodness of fit;
2) the best regression model (denoted by L), representing a simple approach to estimate a given time series without attempting to explicitly connect the response variable to a specific hypothesis about the process that created the data.
Obtained results are shown in Figures 20 21 22 23 24 25. A discussion on model performances obtained is provided in the main text.

	DIC					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	3704	2825	3254	2009	3388	4895
S	3638	2799	3073	1901	3326	4992
$N^{\#}$	3638	2803	3160	2099	3331	4890
$N^{\#\#}$	3610	2710	3104	1906	3286	4806
$N^{\#\#\#}$	3627	2783	3163	2106	3325	4968
$SN^{\#}$	3579	2707	3089	1830	3275	4682
$SN^{\#\#}$	3584	2702	3091	1840	3288	4724
$SN^{\star}$	3546	2668	3043	1830	3269	4682

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 8: DIC values associated with different models and epidemics

	AIC					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	3709.75	2831.26	3259.58	2017.67	3392.09	4907.52
S	3648.48	2808.03	3080.87	1911.98	3369.43	5000.99
$N^{\#}$	3645.38	2816.93	3168.2	2107.09	3337.96	4997.06
$N^{\#\#}$	3618.33	2722.36	3112.32	1924.34	3294.66	4814.95
$N^{\#\#\#}$	3635.1	2792.92	3173.02	2119.44	3333.81	4997.7
$SN^{\#}$	3589.08	2720.32	3103.58	1845.97	3284.13	4692.23
$SN^{\#\#}$	3596.67	2719.83	3103.32	1852.43	3301.37	4736.16
$SN^{\star}$	3556.51	2681.16	3059.46	1853.94	3280.18	4693.88

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 9: AIC values associated with different models and epidemics

	Probability of minimizing the estimated information loss					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
S	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$N^{\#}$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$N^{\#\#}$	< 0.001	< 0.001	< 0.001	< 0.001	0.001	< 0.001
$N^{\#\#\#}$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$SN^{\#}$	< 0.001	< 0.001	< 0.001	-	0.139	-
$SN^{\#\#}$	< 0.001	< 0.001	< 0.001	0.04	< 0.001	< 0.001
$SN^{\star}$	-	-	-	0.019	-	0.438

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 10: Probability of minimizing the estimated information loss associated with different models and epidemics

	BIC					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	3724.74	2849.24	3274.89	2036.04	3405.97	4946.04
S	3669.46	2829.01	3102.31	1933.42	3393.73	5027.96
$N^{\#}$	3663.36	2834.91	3186.57	2125.46	3358.79	5020.17
$N^{\#\#}$	3639.31	2743.34	3133.76	1945.77	3318.97	4841.92
$N^{\#\#\#}$	3656.08	2813.9	3194.46	2140.88	3358.11	5024.66
$SN^{\#}$	3613.05	2744.3	3128.08	1870.47	3311.9	4723.05
$SN^{\#\#}$	3623.64	2746.8	3130.88	1880	3332.62	4770.83
$SN^{\star}$	3583.49	2708.14	3087.03	1881.5	3311.43	4728.55

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 11: BIC values associated with different models and epidemics

	RMSE					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	134415	9729.55	25067.1	1184.8	714.938	287.673
S	516354	8085.85	17189.5	1608.81	788.141	319.61
$N^{\#}$	97865.8	6931.29	11661.5	1012.63	569.433	317.6
$N^{\#\#}$	97785.4	6275.79	18037.8	1091.13	568.696	260.065
$N^{\#\#\#}$	97350.7	25873.7	14151.3	884.04	552.511	316.322
$SN^{\#}$	93953.5	6136.29	18599.7	960.875	602.831	213.666
$SN^{\#\#}$	93082.9	6152.49	17983.7	967.233	572.839	226.179
$SN^{\star}$	110566	4993.79	14379.1	835.451	591.411	213.693

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 12: Root mean square errors (MSE) associated with different models and epidemics

	MAPE					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	84.895	109.63	72.545	203.127	71.238	14.066
S	70.214	97.871	33.393	58.985	74.089	17.231
$N^{\#}$	61.959	88.998	53.319	447.078	59.485	16.8
$N^{\#\#}$	54.039	59.74	40.435	112.926	55.2	12.522
$N^{\#\#\#}$	56.883	80.224	52.509	374.899	57.034	16.772
$SN^{\#}$	45.559	56.808	39.525	55.942	52.657	10.142
$SN^{\#\#}$	46.024	56.264	37.789	53.437	55.747	11.047
$SN^{\star}$	39.597	46.521	30.482	56.774	51.203	10.138

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 13: Mean absolute percentage errors (MAPE) associated with different models and epidemics

	R^2					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	0.352	0.289	0.523	0.492	0.311	0.589
S	-8.853	0.509	0.775	0.041	0.161	0.493
$N^{\#}$	0.657	0.639	0.897	0.628	0.563	0.499
$N^{\#\#}$	0.657	0.704	0.752	0.569	0.564	0.664
$N^{\#\#\#}$	0.659	-4.744	0.846	0.71	0.588	0.503
$SN^{\#}$	0.681	0.717	0.736	0.666	0.506	0.773
$SN^{\#\#}$	0.689	0.715	0.754	0.662	0.558	0.746
$SN^{\star}$	0.551	0.809	0.841	0.746	0.526	0.773

$^{\#}$ based on equation 2 $^{\#\#}$ based on equation 3 $^{\#\#\#}$ based on equation 4
* based on equation 4 and representing our baseline formulation

Table 14: Values of the coefficient of determination (R^2) associated with different models and epidemics

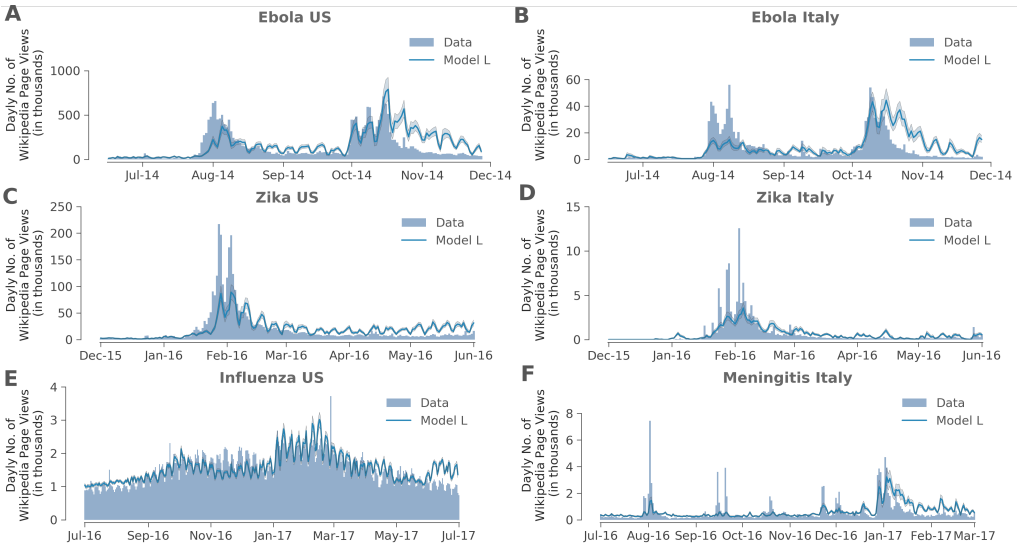

Figure 12: Model estimates based on model L for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

	Pearson correlation coefficient					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	0.666	0.61	0.787	0.782	0.569	0.781
S	0.584	0.729	0.89	0.805	0.524	0.706
$N^{\#}$	0.815	0.807	0.952	0.821	0.819	0.717
$N^{\#\#}$	0.827	0.905	0.925	0.816	0.783	0.816
$N^{\#\#\#}$	0.832	0.706	0.932	0.874	0.792	0.719
$SN^{\#}$	0.839	0.865	0.929	0.82	0.738	0.88
$SN^{\#\#}$	0.84	0.875	0.927	0.818	0.778	0.865
$SN^{\star}$	0.844	0.913	0.929	0.871	0.739	0.88

$^{\#}$ **based on equation 2** $^{\#\#}$ **based on equation 3** $^{\#\#\#}$ **based on equation 4**
* **based on equation 4 and representing our baseline formulation**

Table 15: Pearson correlation coefficients associated with different models and epidemics

	MAE					
Model	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
L	89000.1	5933.26	12204.6	458.193	379.561	213.492
S	149324	4683.11	6575.59	518.305	401.479	252.701
$N^{\#}$	67808.4	4295.07	6329.59	404.018	285.317	242.535
$N^{\#\#}$	63624.1	3544.4	6682.43	356.824	280.194	191.581
$N^{\#\#\#}$	62419.6	6961.49	7496.41	372.234	275.066	242.191
$SN^{\#}$	54976	3410.05	7320.11	346.021	285.435	152.55
$SN^{\#\#}$	54330.9	3370.27	6662.37	343.464	281.787	166.454
$SN^{\star}$	53925.6	2697.62	5983.21	289.98	277.237	152.597

$^{\#}$ **based on equation 2** $^{\#\#}$ **based on equation 3** $^{\#\#\#}$ **based on equation 4**
* **based on equation 4 and representing our baseline formulation**

Table 16: Mean absolute errors associated with different models and epidemics

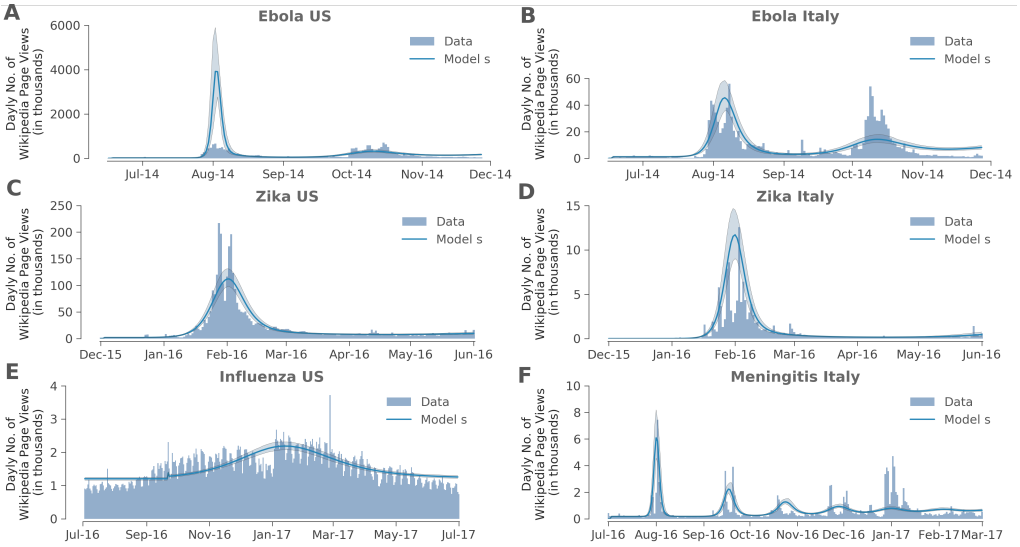

Figure 13: Model estimates based on model S for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

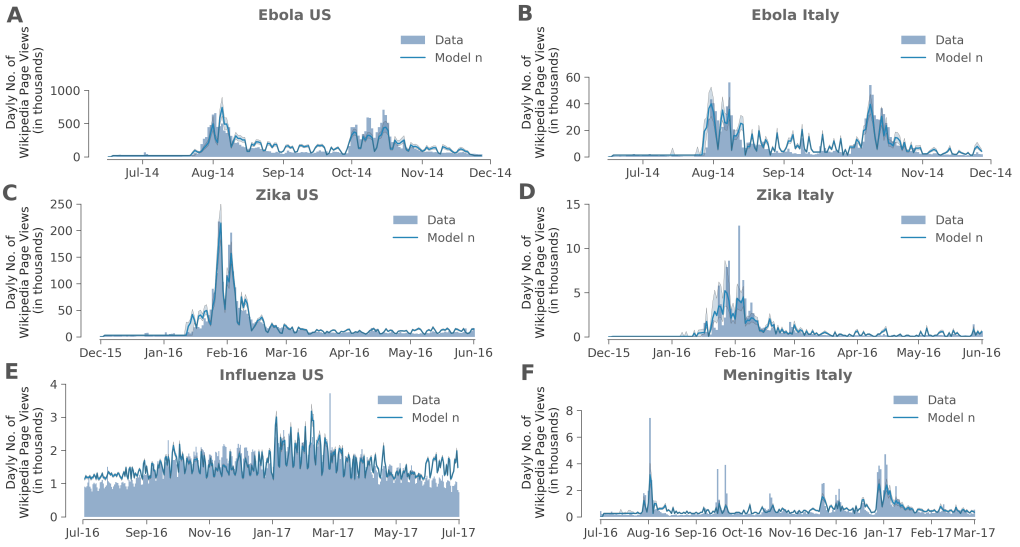

Figure 14: Model estimates based on model N and equation 2 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

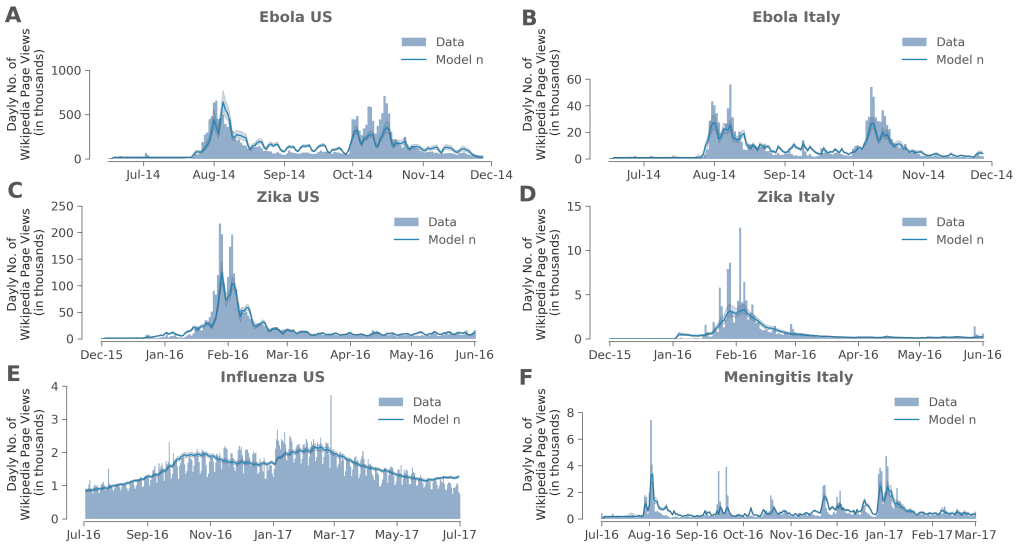

Figure 15: Model estimates based on model N and equation 3 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

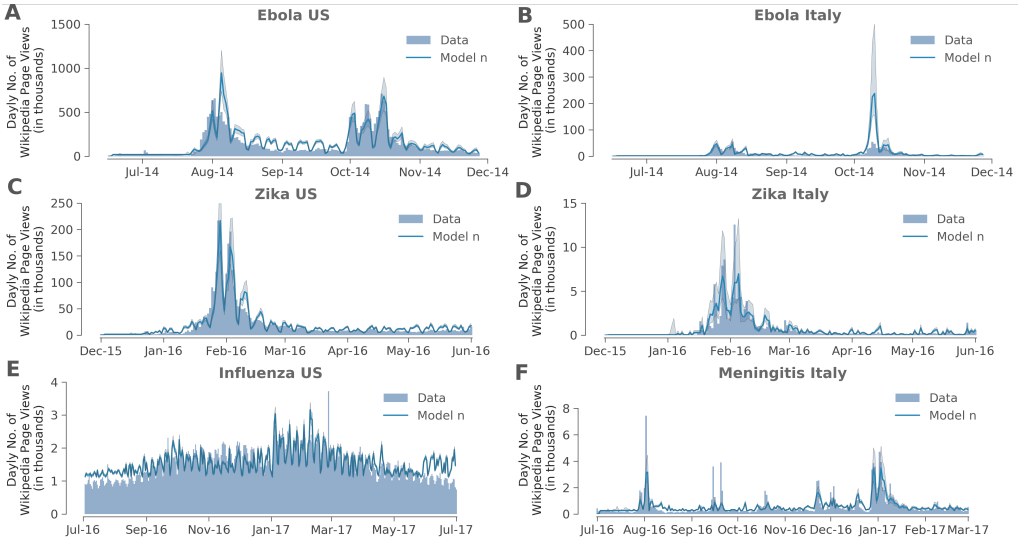

Figure 16: Model estimates based on model N and equation 4 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

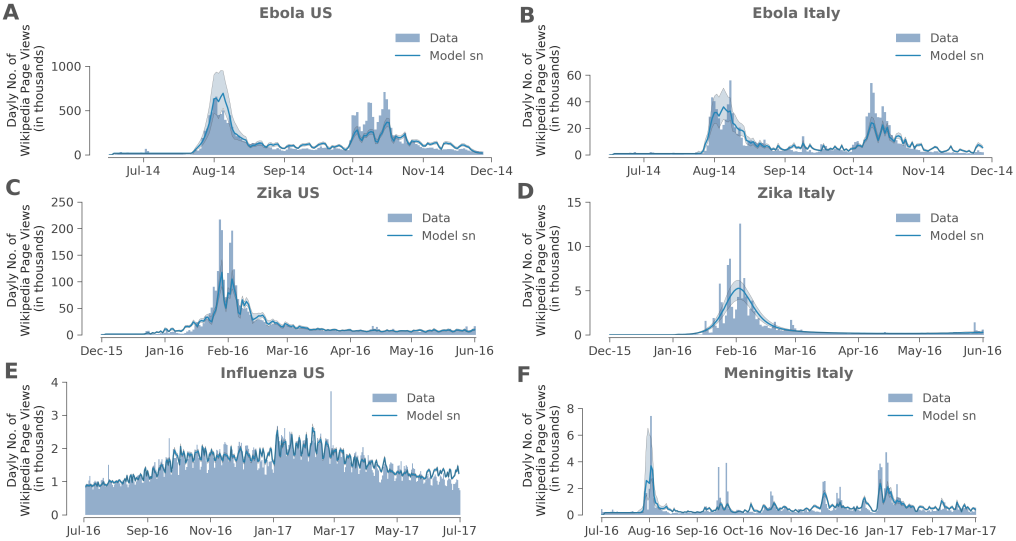

Figure 17: Model estimates based on model *SN* and equation 2 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

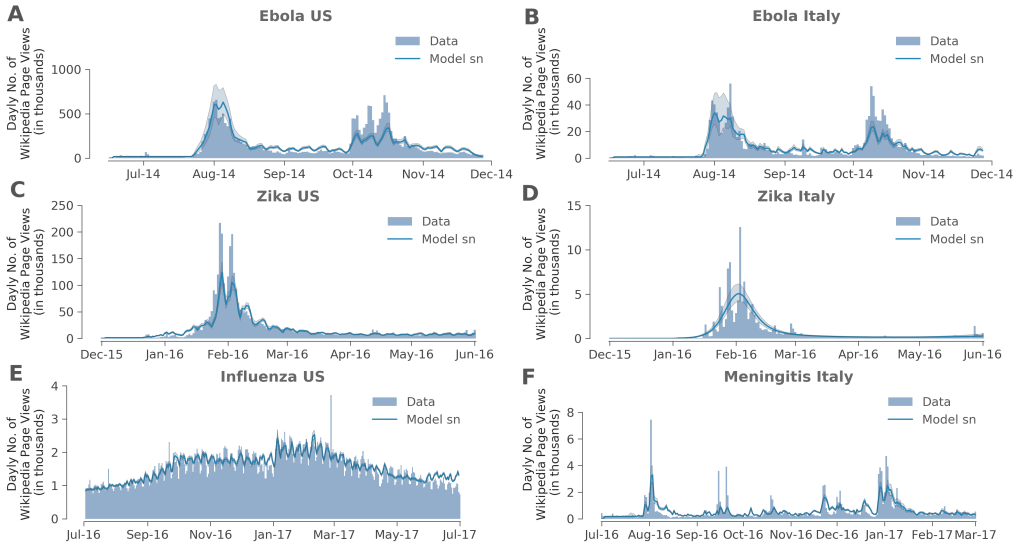

Figure 18: Model estimates based on model *SN* and equation 3 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

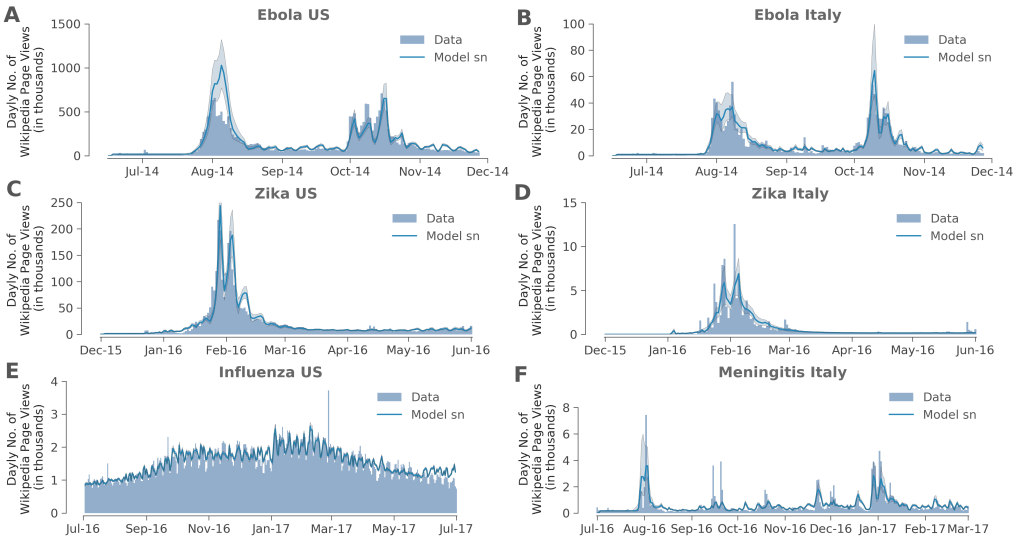

Figure 19: Model estimates based on model *SN* and equation 4 for Ebola **(A)**, Zika **(C)**, Influenza **(E)** in the US and for Ebola **(B)**, Zika **(D)**, Meningitis **(F)** in Italy. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia.

Parameter	**Ebola US**	**Ebola IT**	**Zika US**	**Zika IT**	**Meningitis IT**	**Influenza US**
k	0.047	0.014	0.0092	0.0018	0.00083	0.0014
β_S	1.19	0.52	0.085	0.11	5.35	0.049
$1/\gamma$	1.21	2.41	43.04	59.41	0.22	20.20
β_N	30579	47570	53797	21528	40763	20368
ρ	3.53	1.79	3.14	1.41	4.96	15.78
$1/\nu$	80.66	113.47	211.44	298.63	53.13	366.41
x	5.57	4.76	7.47	3.85	3.43	59.85
W_0	17396	956	1471	2	166	847
t_0	22/07/14	23/07/14	22/12/15	1/01/16	26/07/16	1/07/16

Table 17: Mean estimates of free model parameters obtained for the SN-model and equation 4

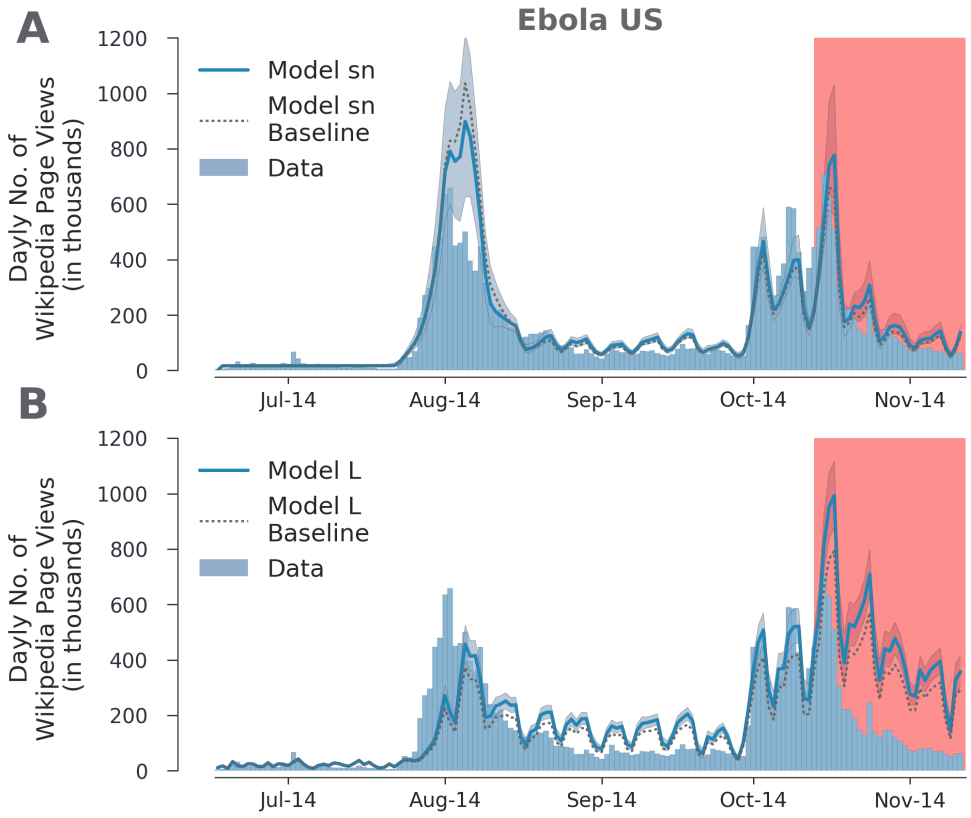

Figure 20: Model estimates obtained with model SN **(A)** and model L **(B)** for Ebola in the US when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

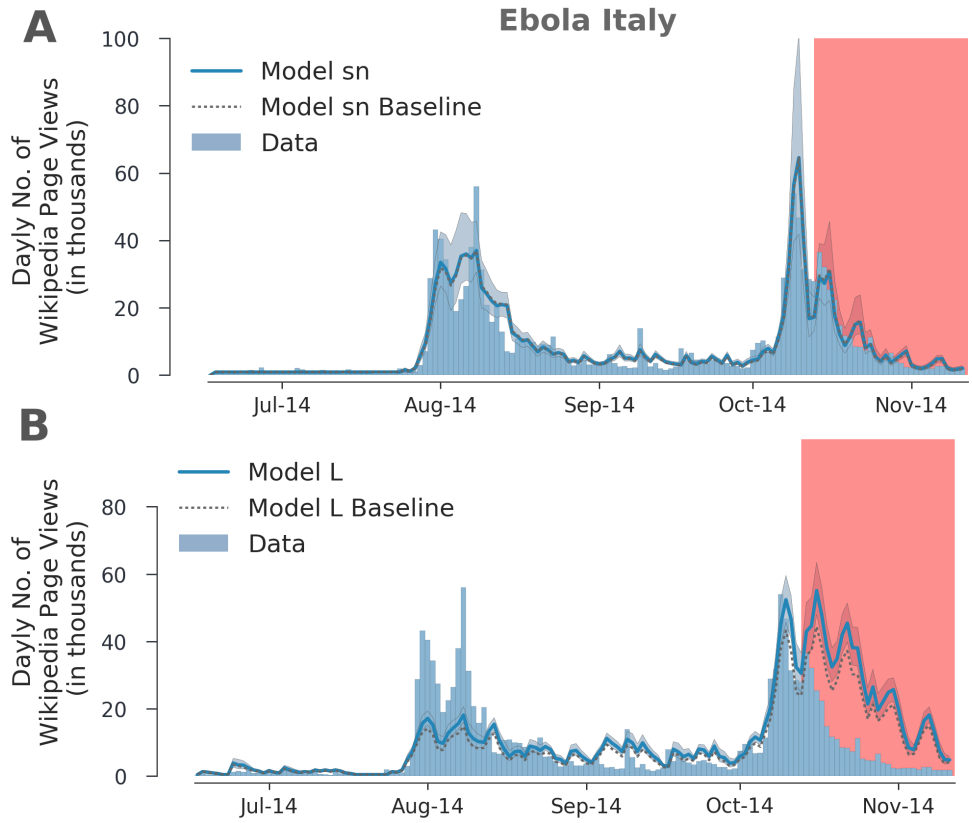

Figure 21: Model estimates obtained with model SN **(A)** and model L **(B)** for Ebola in the Italy when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

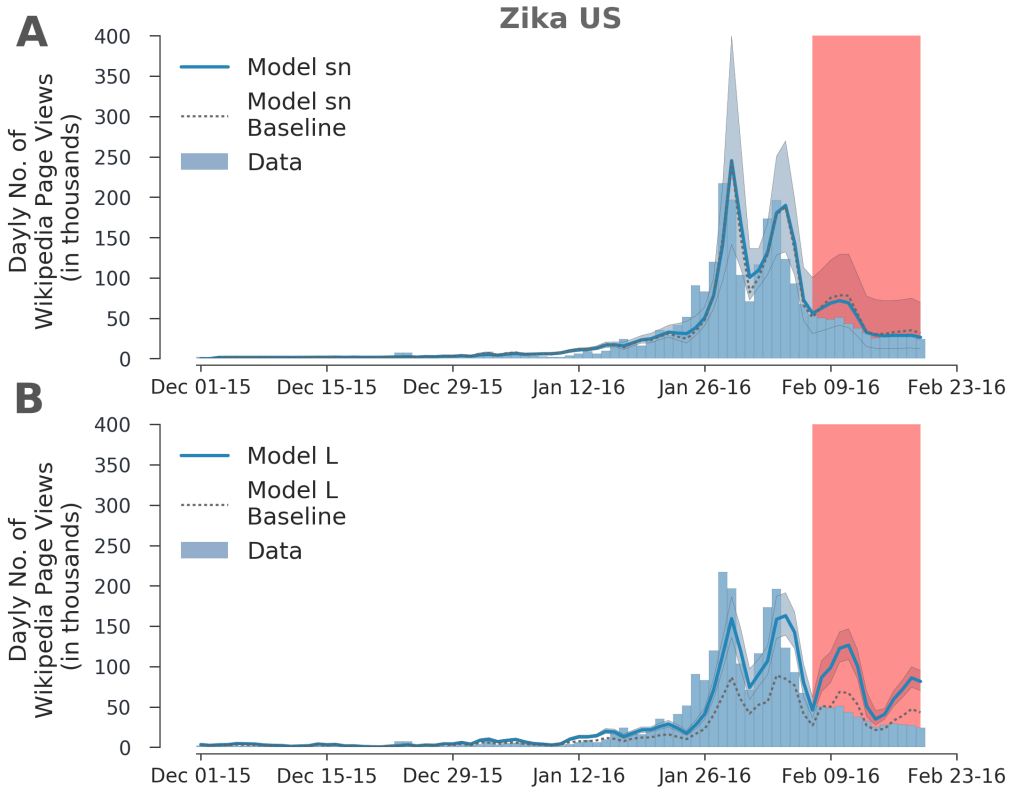

Figure 22: Model estimates obtained with model SN **(A)** and model L **(B)** for Zika in the US when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

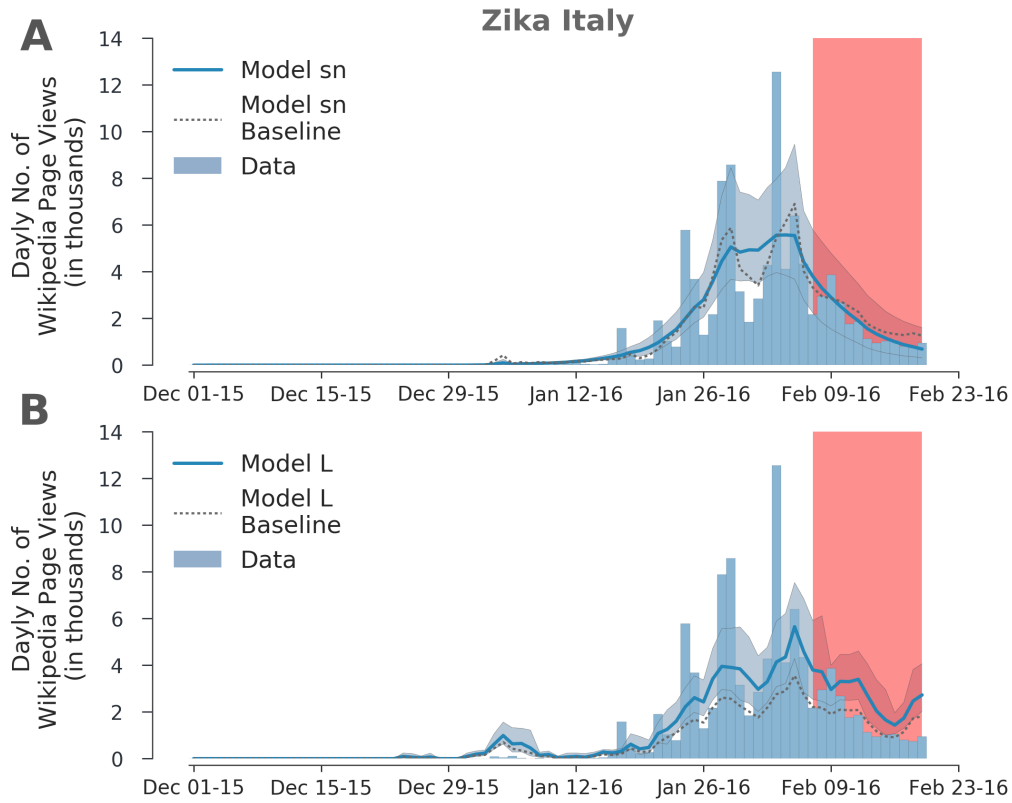

Figure 23: Model estimates obtained with model SN **(A)** and model L **(B)** for Zika in Italy when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

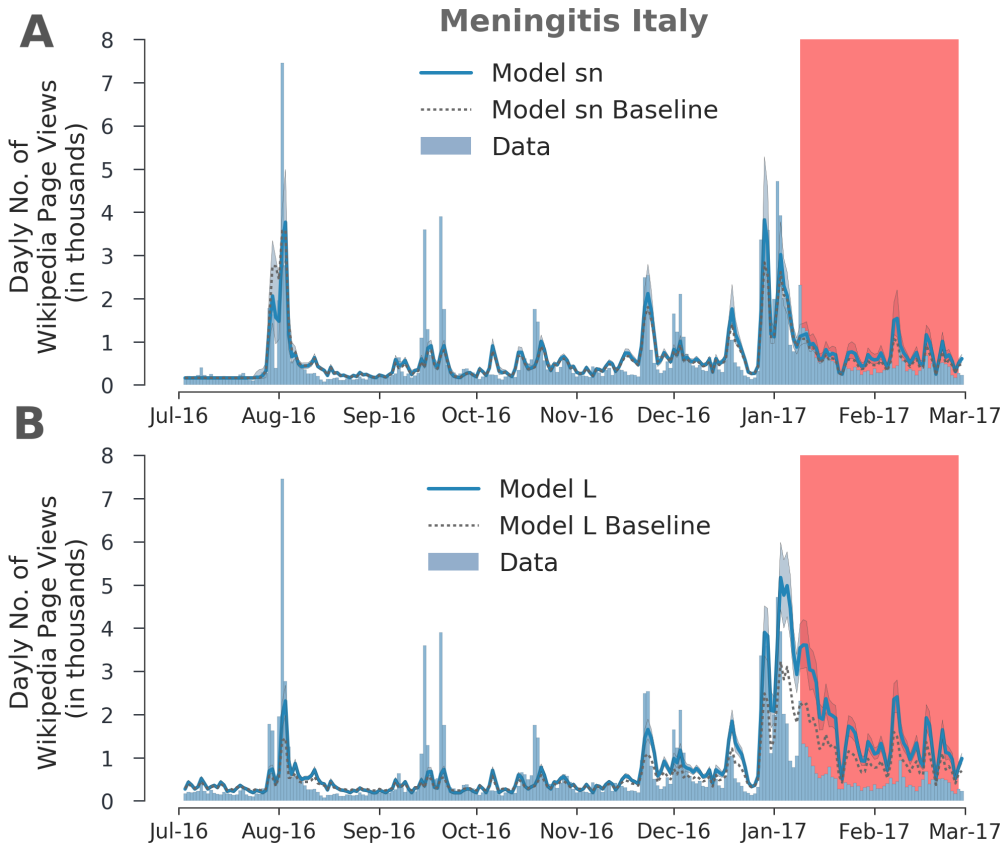

Figure 24: Model estimates obtained with model SN **(A)** and model L **(B)** for Meningitis in Italy when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

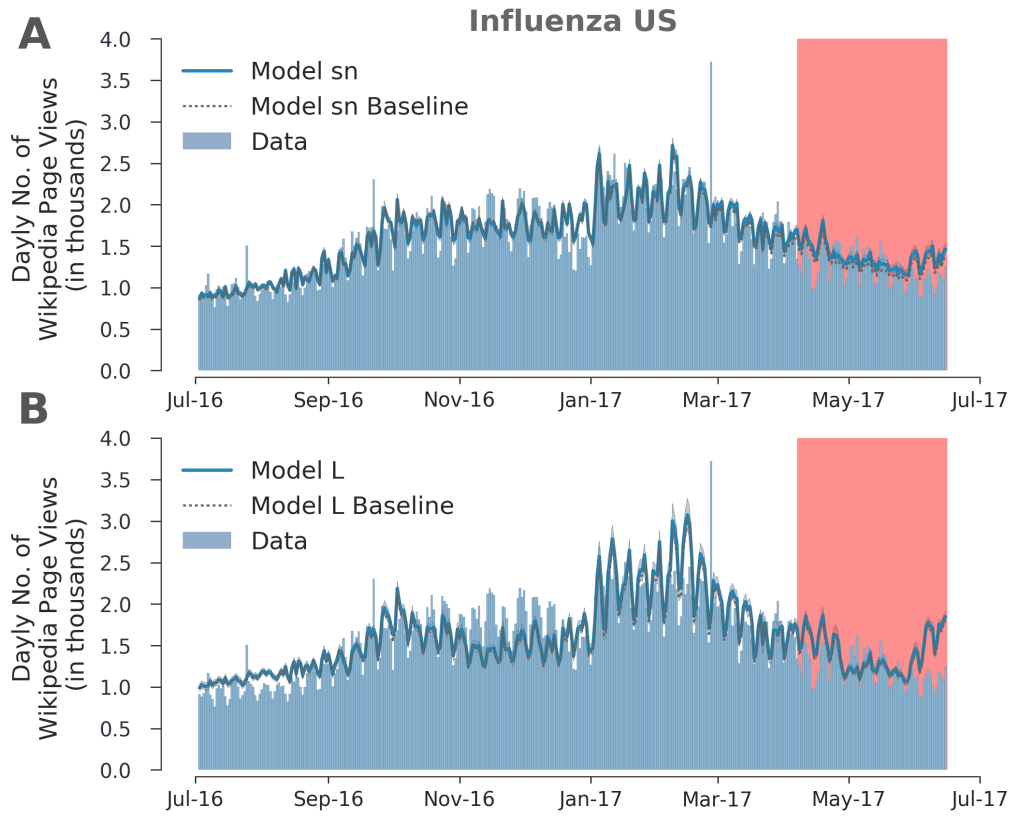

Figure 25: Model estimates obtained with model SN **(A)** and model L **(B)** for Influenza in the US when only the first 80% of data points were used for model calibration. In each panel, blue bars represent the daily number of Wikipedia page views over time for the considered infection. The blue lines and the shaded areas refer to the average and the 95% CI of estimates as obtained with the model on the daily number of informed individuals seeking information on Wikipedia. Red regions highlight the 20% of data points that has not been used for model calibration. The dashed grey lines show model estimates as obtained when using 100% of data points for model calibration (baseline).

3

REDUCING MEASLES RISK IN TURKEY THROUGH SOCIAL INTEGRATION OF SYRIAN REFUGEES

3.1 INTRODUCTION

Human migration represents a complex phenomenon influencing in several interconnected ways the economy, the healthcare and the social cohesion of whole countries [3, 140, 29, 173, 26, 126, 18]. However, it is only recently that the availability of massive datasets opened to new advancements in modelling and understanding such complexity [96, 38, 19, 97], addressing the urgent need for effective, large-scale intervention policies towards managing the consequences of massive migration flows [26].

In this chapter, we focus our attention on Turkey, a country facing a humanitarian emergency of unprecedented levels [152]. In the last eight years, more than 3.5M Syrians, displaced by the war, have sought refuge in Turkey. This number, through births and new arrivals, is also increasing by approximately 1,000 people per day. The arrival of a huge amount of people with different economic, health, and living conditions, and from a country where the healthcare system has been almost completely disrupted, may raise serious concerns about the risks of Turkish health systems being overburdened.

For instance, Turkish infectious disease specialists are concerned that Syrian refugees' crisis may impose serious risks to their country for infectious diseases previously eliminated or in the process of being eliminated [73]. According to the latest reports from WHO and UNICEF [166], immunization coverage in Syria dropped from more than 80% before the war to a worrying 41% in 2015 for the most basic vaccines, resulting in millions of unvaccinated children. Direct consequences of this alarming situation are a high risk of epidemic outbreaks (e.g., evidence for polio [167] and measles [164] has been reported) and a potential increase of mortality due to diseases which could be prevented with vaccines [138]. Thus, countries, such as Turkey, Lebanon, and Jordan, hosting a great concentration of Syrians perceive the lack of an appropriate immunization coverage as a potential risk of epidemic outbreaks for the local population [73]. This perceived risk may ignite a cascade of social dynamics which could: i) reinforce the segregation of refugees; ii) increase unemployment and poverty; iii) result in difficult relationships between healthcare workers and Syrians.

The aim of this study is to quantify the risk of observing widespread measles epidemics in Turkey, showing potential public health benefits coming from social in-

tegration between Syrian refugees and Turkish citizens. To such aim we developed a transmission model to investigate the influence of social mixing and integration among Syrian refugees and Turkish citizens on the potential spread of measles epidemics in Turkey. The model takes explicitly into account empirical mobility patterns, as inferred by mobile phone data [129], and the current level of immunity against measles in the two considered populations, as estimated from available epidemiological evidences [165, 166]. Since the current amount of integration is difficult to estimate with available data, measles transmission is modelled by considering a tunable parameter that accounts for a variety of scenarios, ranging from full segregation to full integration. Results are obtained by simulating the spread of measles by assuming different scenarios for measles transmission potential and different levels of social integration.

In particular, measles represents an illustrative case of a highly contagious infectious disease which can be prevented with a safe and effective vaccine [69, 90, 138, 106, 91, 151, 124]. Despite substantial progress towards measles elimination at the global level has been documented, re-emergence of large measles epidemics was observed in the last decade both in low-income and in high-income countries [151]. Moreover, measles epidemiology varies widely across different geographical regions, as a consequence of heterogeneous immunity gaps, generated by sub-optimal immunization activities, in different socio-demographic settings [106, 151, 91].

The crucial role played by both human mobility [48, 32, 10, 105, 108, 171] and mixing patterns [111, 54, 5] in shaping the transmission dynamics of infectious diseases has been widely documented in the literature and represents a key component of realistic modeling aimed at informing public health policies. Thus, human mobility models have been used to map flows of individuals between geographical areas at different scales and to improve the reliability of transmission models of infectious diseases [48, 32, 10, 104, 9, 93, 66]. In the last years, mobile phone data have been successfully used as a valuable proxy for human mobility [67, 142, 137, 149, 19, 14]. We capitalize on these works to build, from mobile phone data, a multilayer network [36, 83, 35] map of human mobility of Turkish citizens and Syrian refugees in Turkey, and we use this knowledge to develop a computational model for the potential epidemic spread of measles.

The contribution of our work is twofold. On the one hand, we identify the *epidemic risks* associated with measles in Turkey. On the other hand, we investigate how these epidemic risks are affected by policies devised to enhance *social integration* between Syrian and Turkish populations.

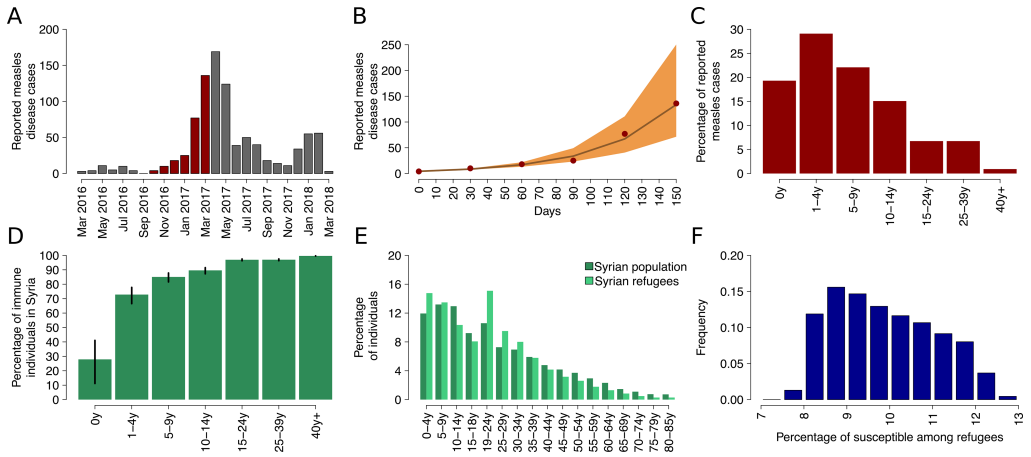

Figure 26: Measles immunity levels. **A)** Reported number of measles disease cases over time, during the 2016-2018 measles epidemics in Syria as recently reported by the World Heath Organization [165]; red bars correspond to data points used to derive the R_e as a function of the exponential growth rate of the observed epidemic. **B)** Obtained fit of the epidemic exponential growth between September and February 2017 in Syria: red solid line represents the mean estimate, orange shaded area represents 95%CI. **C)** Observed distribution of measles cases across different ages during the 2016-2018 measles epidemics in Syria as recently reported by the World Heath Organization [165]. **D)** Estimated age specific serological profile in Syria at the beginning of 2017: green bars represents mean values, vertical black lines represent 95%CI. **E)** Observed age distribution of Syrian refugees in Turkey (light green) [1] compared with the population age distribution in Syria (dark green). **F)** Estimated percentage of susceptible individuals among Syrian refugees.

3.2 RESULTS

3.2.1 *Immunity levels in the two populations.*

Two different immunity levels against measles infection are estimated for the Turkish and the Syrian populations. As measles epidemics have not been recently reported in Turkey, we assume the measles immunity level among Turkish citizen reflects the fraction of immunized individuals among birth cohorts between 2006-2016 through 1^{st} and 2^{nd} dose routine vaccination programs [166] (see Appendix). Accordingly, our estimates suggest that only 3.8% of Turkish people might be currently susceptible to measles infection. A different level of susceptibility in the Turkish population is also considered for sensitivity analysis in the Appendix section of this chapter.

Estimates of the immunity level among refugees was instead obtained by inferring the age-specific fraction of susceptible individuals in Syria during a recent measles epidemic from the growth rate and age-distribution of cases reported in 2017 [165], and accounting for the age distribution of Syrian refugees in Turkey [1] (Fig. 26). We found that the effective reproductive number (R_e) of the recent Syrian measles epi-

demic was 1.32 (95%CI 1.26-1.38). Consequently, we estimated that the percentage of susceptible individuals in Syria at the beginning of 2017 was 8.92% (95%CI 7.29-10.96). The resulting percentage of susceptible individuals among Syrian refugees in Turkey was estimated to be 9.87% (95%CI 8.07-–12.18)(Fig. 26).

Obtained results suggest that nowadays, in Turkey, 280,000-430,000 out of 3.5M Syrian refugees and about 3M out of 80M Turkish people are measles susceptible.

3.2.2 *Social integration, human mobility and disease transmission.*

The risk of measles re-emergence in Turkey is here analyzed by using a compartmental transmission model explicitly taking into account potential infectious contacts occurring between individuals moving across the country. Different scenarios on how much Syrian refugees interact with Turkish citizens are investigated. A schematic representation of the model is shown in Fig. 27 A,B along with spatial mobility patterns inferred by the analysis of Call Detailed Records (CDRs) associated with the usage of mobile phones in the country.

The fundamental quantity regulating disease dynamics is the basic reproduction number (R_0), which represents the average number of secondary infections in a fully susceptible population generated by a typical index case during the entire period of infectiousness. Larger R_0, higher the disease transmissibility. If $R_0 > 1$ the infection will be able to spread in a population. Otherwise, the infection will die out. For endemic diseases like measles, R_0 provides insights into the proportion p of immune population (either due to vaccination or natural infection) required to prevent large outbreaks; the equation $p = 1 - 1/R_0$ represent the proportion of immune population needed to prevent the spread of the disease as a function of the basic reproduction number R_0 [6, 69, 90]. For instance, if $R_0 = 20$ at least 95% of the population has to be immune to eliminate the disease. As for measles, typical values of R_0 ranges from 12 to 18 [6, 69, 90, 106, 124]. However, when considering diseases with pre-existing levels of immunity (e.g. childhood diseases like measles), R_0 is a theoretical value representing what could happen in terms of disease transmissibility by removing immunity. In these cases, an appropriate measure of diseases transmissibily is provided by the effective reproduction number (R_e), which represents the average number of secondary infections in a partly immunized population generated by a typical index case during the entire period of infectiousness.

In Fig. 27 C we show the ratio R_e/R_0 as obtained by varying the fraction of Syrian refugees susceptible to measles from 8.07% to 12.18% and by varying the level of social integration from 0% (full segregation of refugees) to 100% (full integration of refugees). We found that pre-existing levels of immunity of the two populations reduce R_e to values lower than 10% of R_0. For example, if R_0 is lower than 10, the probability of observing an epidemic outbreak would be close to 0 because R_e would result

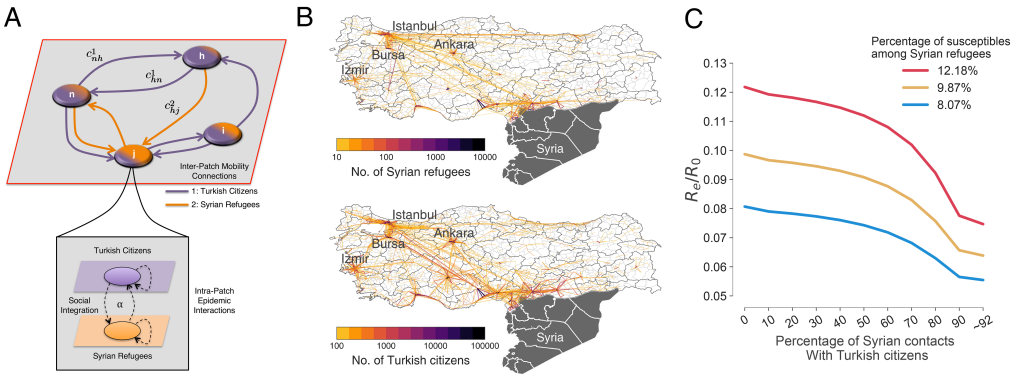

Figure 27: Model structure, human mobility and transmission potential. **A)** Schematic illustration of the model considered in this work. Each prefecture of Turkey is considered as a node of a meta–population network of geographic patches. Two populations, namely Turkish and Syrians, are encoded by different colors and move between patches following the inferred inter-patch mobility pathways. Turkish and Syrian populations encode two different layers of a multilayer system [36, 83, 35] where social dynamics and epidemics spreading happen simultaneously. **B)** Mobility of Syrian refugees (Top) and Turkish citizens (Bottom) between the prefectures of Turkey as inferred from CDR. Different colors are used to indicate the number of individuals moving from a prefecture to another. **C)** Effective reproduction number for measles spreading according to our model, rescaled by R_0, as a function of the mixing parameter accounting for social integration between Turkish and Refugees. Colored lines are associated with the estimated levels of susceptibility among Syrian refugees.

lower than 1 as a consequence of pre-existing immunity levels. However, if R_0 is in a more plausible range of values (e.g. 12-18), pre-existing levels of immunity, which are particularly low among Syrian refugees, might not be sufficient to prevent the spread of the disease. Moreover, we found that R_e is maximum when the two populations live socially segregated from each other, whereas it quickly decreases by almost 50% when the two populations are socially well integrated.

In sum, the immunity level characterizing the Turkish population in 2017 is expected to prevent the spread of future measles epidemic in geographical locations predominantly populated by Turkish citizens. However, if a measles index case would occur in a population with a sufficiently large proportion of Syrian people, transmission events will be sustained by the lack of adequate immunity levels among refugees. Our modelling analysis show that for any scenario considered the risk of observing large epidemics increases with the basic reproduction number and the proportion of susceptible among the refugees (see Fig. 28 and Appendix section).

3.2.3 *Measles epidemic risks in Turkey.*

In case of full segregation of refugees (although practically infeasible and therefore unlikely), potential measles epidemics would result in dramatic health consequences among refugees, causing a huge amount of measles cases widespread in the country (Fig. 28). Specifically, when $R_0 = 15$ is considered and 9.8% of refugees are assumed to be measles susceptible, the probability of observing an epidemic with more than 20 cases is 100% (see Supplementary Discussion) and the final size of potential epidemics is expected to exceed 10,000 cases (mean estimate 10,662 95%CI 3,172–18,414, see Fig. 28). Our results show that the risk of observing sustained transmission in the country is large for any value of R_0 larger than 15 but also for lower values of R_0 (e.g. $R_0 = 12$) if the proportion of refugees susceptible is 9.8% or more (see Fig. 28 and Appendix).

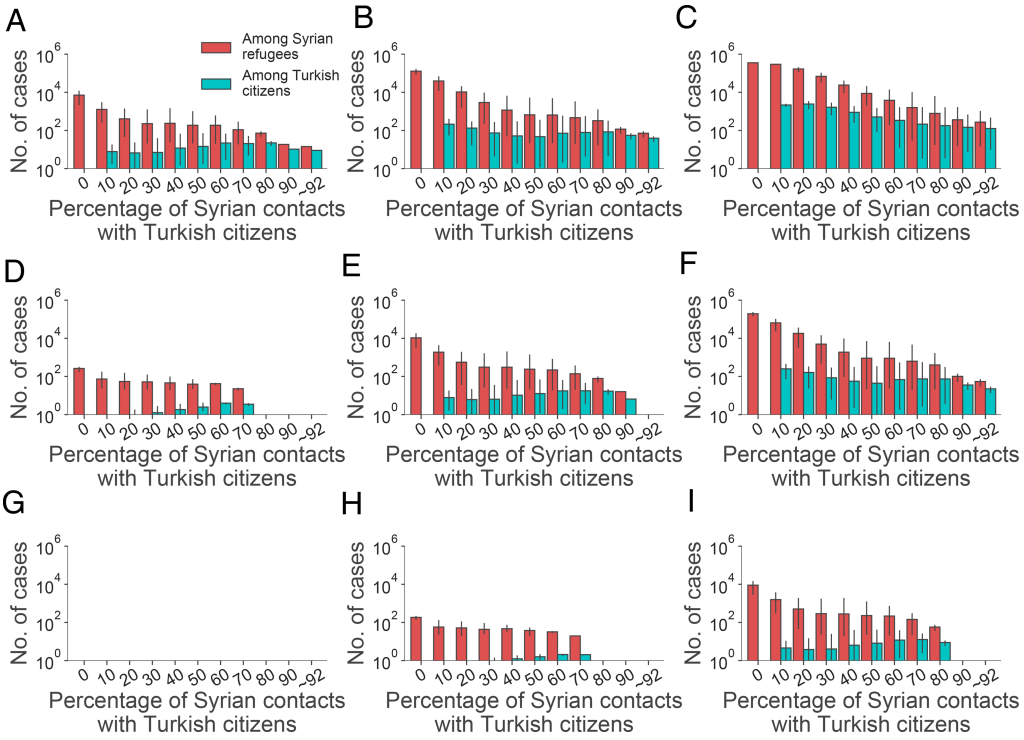

Figure 28: Cumulative infections considering epidemics that exceed 20 cases in the entire population. Bars represent the average number of infections occurring among Syrian refugees (red) and Turkish citizens (blue) for the model projections as a function of the mixing parameter, black lines represent 95%CI. Three different values of R_0 and Susceptibility levels among the Syrian refugees population (S_0^{Ref}) were considered. **A)** $R_0 = 18$, $S_0^{Ref} = 8.07\%$. **B)** as A but for $S_0^{Ref} = 9.87\%$. **C)** as A but for $S_0^{Ref} = 12.15\%$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

In the case of full segregation, infections would occur only among refugees. However, when assuming high level of segregation (i.e. only a small fraction, yet equal or greater than 10%, of refugees' contacts occur with Turkish citizens), the risk of experiencing large measles outbreak is high (see Appendix) and measles epidemics could produce non-negligible spillover of cases among Turkish citizens as well (Fig. 28). In particular, in a worst case scenario where $R_0 = 18$, 12.18% of refugees are susceptible and more than 70% of Syrian contacts occur with Syrian people, thousands of measles cases are expected all over the country among the Turkish people as well (see Fig. 28, 29).

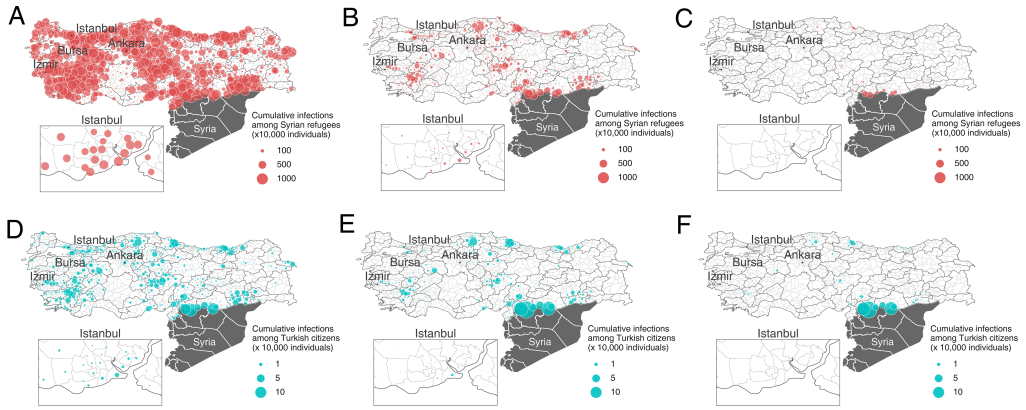

Figure 29: The potential spatial burden of epidemics. **A)** shows the estimated cumulative infections in the case of 20% of Syrian contacts with Turkish citizens considering the worst case scenario in terms of R_0 and immunity levels against measles infection among Syrian refugees. Bubbles size are proportional to the average number of measles cases in the Turkish prefectures per 10,000 individuals. Inset displays the Istanbul prefectures. **B)** as A but for 40% of Syrian contacts with Turkish citizens. **C)** as A but for 60% of Syrian contacts with Turkish citizens. **D)** as A with respect to the Turkish population. **E)** as B with respect to the Turkish population. **F)** as C with respect to the Turkish population.

More in general, obtained results suggest that the risk of observing sustained measles transmission, the final size of potential epidemics and the populated area at risk of measles infection are significantly smaller in the presence of high levels of integration of refugees (Fig. 28, 29). Specifically, when $R_0 = 15$ is considered, 9.8% of refugees are assumed to be measles susceptible and refugees well mix with the Turkish (e.g. more than 70% of Syrian contacts occur with Turkish people), the probability of observing epidemic outbreak dramatically decreases to values lower than 10% (see Appendix section). Moreover, in case of outbreak, the expected overall number of cases is no larger than few hundred (Fig. 28), as potentially infectious contacts would more probably occur with Turkish immune individuals, who represent about 90% of individuals currently leaving in Turkey.

3.2.4 *Spatial diffusion of potential epidemics.*

Remarkably, larger segregation levels also promote the spatial invasion of the epidemic across the whole country (Fig. 30). In the worst case scenario of $R_0 = 18$, 12.18% of Syrian refugees susceptible, and more than 90% of Syrian contacts occur within the Syrian population, the measles epidemic is expected to affect more than 300 out of 1021 prefectures of Turkey (Fig. 30 A). On the opposite, if more than 70% of contacts of refugees would occur with Turkish people, as a consequence of good integration of refugees with Turkish citizens, for the majority of epidemiological scenarios considered, measles epidemics are expected to remain geographically bounded in less than 10 prefectures of the country (Fig. 30 and Appendix).

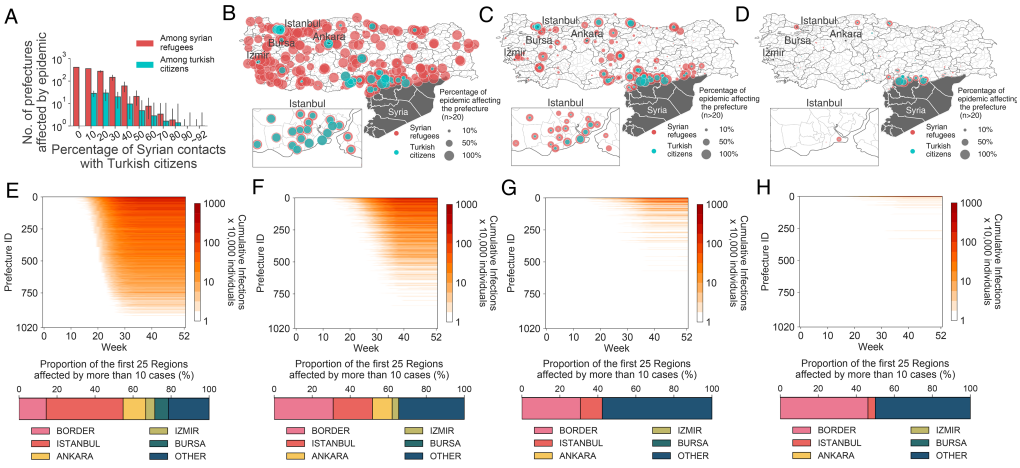

Figure 30: Spatio-temporal spread of potential epidemics. **A)** shows the estimated number of prefectures affected by the epidemic as a function of the mixing parameter in the worst case scenario. Bars represent the average number of prefectures exceeding 20 cases among Syrian refugees (red) and Turkish citizens (blue); black lines indicate the 95%CI. **B)** Percentage of the simulated epidemic that exceed 20 cases per prefecture in the case of 20% of Syrian contacts with Turkish citizens. Red and blue bubbles refer to Syrian refugees and Turkish citizens respectively. **C)** as B but for 40% of Syrian contacts with Turkish citizens. **D)** as B but for 60% of Syrian contacts with Turkish citizens. **E)** (Top) relative incidence over time considering both the populations per prefecture in the case of total segregation. Prefectures are ranked in decreasing order at week 20. (Bottom) Proportion of region affected in the initial phase of the epidemic considering the first 25 prefectures affected by more than 10 cases. Border refers to Hatay, Kilis, Gaziantep, Sanliurfa, Mardin and Sirnak regions. **F)** as E in the case of 20% of Syrian contacts with Turkish citizens. **G)** as E in the case of 40% of Syrian contacts with Turkish citizens. **H)** as E in the case of 60% of Syrian contacts with Turkish citizens.

Furthermore, our results suggest that the level of social integration between refugees and the Turkish population can also strongly affect the spatio-temporal spread of potential measles epidemics. Figure 30 E-H shows for each prefecture the expected cu-

mulative measles incidence over time for different levels of social integration in the the worst case scenario. Obtained estimates indicate that, in the case of full segregation, 57% of Turkish prefectures is expected to experience more than 10 measles cases after 30 weeks since the beginning of an epidemic. Such percentage decreases to 11%, 1% and 0.4% when refugees contacts with Turkish citizens increases to 20%, 40% and 60% respectively. Interestingly, in case of full segregation, prefectures where cases of infections are registered at earlier stages are mainly located in regions associated with the four largest cities of Turkey (64%, see Figure 30 E). In contrast, when 60% of refugees' contacts occur with Turkish citizens, a remarkable fraction of prefectures that would be affected the early phase of a measles epidemic are close to the Syrian border (Figure 30 H). This partially explains why social integration can - during a potential measles epidemic - significantly reduce the spillover of cases in the Turkish population.

3.3 DISCUSSION

The widely accepted critical immunity threshold for measles elimination is 95% of immune individuals. According to our estimates, while Turkish citizens are mostly protected by high vaccine uptake levels, Syrian refugees display a considerably larger fraction of individuals that is susceptibile to the measles, as a consequence of the suboptimal vaccination during the ongoing civil war. More specifically, while the level of protection of the Turkish population against the disease is nearly optimal (more than 96% of immune individuals), the protection of Syrian refugees is far from being acceptable (only about 90% of immune individuals, though highly uncertain).

The strong difference in the immunity levels among the two populations may have deep repercussions on the society perception towards the movement of Syrians within Turkey. As common in Western countries hosting considerable amounts of migrants [34], Turkish citizens might perceive the lower immunization coverage of Syrian refugees as a potential threat to national welfare and health. This perception might be even worsened by the staggering numbers of Syrian refugees registered in Turkey, 3.5M in 2018 [129, 34]. This well documented negative perception [43, 13] may trigger segregation mechanisms, aimed at reducing as much as possible interactions and contacts between Syrian refugees and Turkish citizens.

The carried out analysis provides compelling evidence that social segregation does not hamper but rather boosts potential outbreaks of measles to a greater extent in Syrian refugees but also in Turkish citizens, although to a lesser extent. The main result of the current study is the quantitative evidence that social mixing among Syrian refugees and Turkish citizens can be highly beneficial in drastically reducing the incidence and the strength of infection of measles. This is due to the fact that the high immunization coverage of Turkish citizens can shield Syrian refugees from getting

exposed to the infection and this reduces potential sources of infection, in a virtuous cycle reminiscent of herd immunity and well documented in many real-world social systems [49, 138]. Our quantitative model combines CDRs data with available epidemiological evidences to estimate the spatial distribution, the immunity profile and the mobility patterns characterizing the two considered populations, allowing the investigation of spatio-temporal patterns of a potential measles epidemic in Turkey.

Provided that a full homogeneous mixing of refugees and citizens could prove to be impracticable or rather difficult to achieve, there are several policies that could reduce social segregation. For instance, designing specific housing policies for redistributing refugees across different neighbourhoods of a given metropolitan area could avoid the creation of ghettos [26], while also increasing the chances of social interactions between refugees and citizens in schools, shops, third places, etc. Although the proposed analysis clearly shows that increasing social mixing between Syrian refugees and Turkish citizens is expected to produce positive public health outcomes, social integration is also expected to provide major societal benefits such as the reduction of violent crimes, economic and educational inequalities [114].

From a geographic perspective, our analysis confirmed that there are metropolitan areas that are pivotal in diffusing the incidence of the disease over time. These areas are mainly prefectures of Istanbul and Ankara and, unsurprisingly, include also many areas adjacent to the national borders of Turkey with Syria. It is in these areas that the efforts for reducing social segregation should be strategically focused. This poses a great challenge for the future, provided that recent reviews of urban regeneration projects highlighted an important process of social segregation of minorities and non-Turkish ethnicity particularly strong in large cities such as Istanbul [43]. However, our results also suggest that social integration would decrease the relevance of large Turkish cities in promoting the spread of the infection. On the other hand, immunization campaigns targeting areas characterized by a large amount of refugees with respect to the Turkish population, as it is the case of many prefectures close to the Syrian border, might critically reduce the chances of measles transmission and prevent the onset of widespread epidemics.

The performed analysis has several limitations that should be considered in interpreting the results. Estimates of immunity levels in Syrian refugees and in Turkish citizens should be considered cautiously as no recent serological surveys are available for the two populations. Immunity levels are inferred from the analysis of vaccine coverage for Turkish citizens and from the analysis of the 2016-2018 measles outbreak in Syria. This last analysis in particular might be affected by under-reporting of cases [138, 124] and does not consider potential spatial heterogeneities that could drastically affect estimates of the overall level of protection against the disease. Also, we assume the same levels of immunity in all municipalities, thus neglecting spatial

heterogeneities that may be present as a consequence of different vaccine uptake across different regions [148]. Moreover, no data on mixing patterns (e.g. by age) are available for either Syrian refugees and Turkish citizens. Consequently, the model neglects potential differences in measles transmissibility by age of individuals and, similarly, potential differences in measles transmissibility for Syrian refugees and Turkish citizens (e.g. induced by different numbers of overall contacts). Finally, CDRs data used in the proposed analysis are associated with only a fraction of the population. Although these data may not perfectly reflect real movements occurring across all the prefectures in the country, they provide valuable evidence to infer a fair approximation of human mobility in the country driving the spatio-temporal spread of an epidemic.

All this considered, the analysis carried out represents a first attempt to quantify the risk of measles outbreak in Turkey and provides striking evidence that, besides policies aimed at increasing vaccination coverage among Syrian refugees, social integration of refugees within the Turkish population might be an effective countermeasure.

3.4 MATERIAL AND METHODS

The epidemic model used in our study is an standard one based on three compartments, namely Susceptible, Infected and Recovered (SIR). We refer to the Appendix section for the details about the model. Here, we provide information about the force of infection used in our model.

To model the mobility of Turkish and Syrian refugees, we assume two populations of individuals, namely population 1 of size $N^{(1)}$ and population 2 of size $N^{(2)}$, living in a territory consisting of L distinct geographically patches (i.e., Turkish prefectures) accounting for $N_k^{(1)}$ and $N_k^{(2)}$ individuals, $k = 1, ..., L$ with $\sum_{k=1}^{L} N_k^{(1)} = N^{(1)}$ and $\sum_{k=1}^{L} N_k^{(2)} = N^{(2)}$.

The absolute number of individuals moving between patches is inferred from available Call Detail Records as in Refs. [93, 100] and rescaled to adequately represent the volumes corresponding to 80M Turkish individuals and 3.5M Syrian refugees.

Let us indicate by $c_{ki}^{(p)}$ $(p = 1, 2)$ the elements of a matrix $\mathbf{C}^{(p)}$ encoding the number of people belonging to population p travelling from patch k to patch i, and with α the fraction of Syrian contacts with Turkish citizens. The force of infection for each population in the i^{th} patch depends on the contribution of all patches in the country:

$$\lambda_i^{(1)}(\alpha, \mathbf{C}^{(1)}, \mathbf{C}^{(2)}) = \beta_1 \sum_{k=1}^{L} \left[\underbrace{c_{ki}^{(1)} \frac{I_k^{(1)}}{N_k^{(1)}}}_{\text{Endogenous}} + \underbrace{\alpha c_{ki}^{(2)} \frac{I_k^{(2)}}{N_k^{(2)}}}_{\text{Exogenous}} \right]$$

$$\lambda_i^{(2)}(\alpha, \mathbf{C}^{(1)}, \mathbf{C}^{(2)}) = \beta_2 \sum_{k=1}^{L} \left[\underbrace{\alpha c_{ki}^{(1)} \frac{I_k^{(1)}}{N_k^{(1)}}}_{\text{Exogenous}} + \underbrace{c_{ki}^{(2)} \frac{I_k^{(2)}}{N_k^{(2)}}}_{\text{Endogenous}} \right],$$

where $\beta_p = \beta / P_i^{(p)}(\alpha, c)$ is the transmission rate for population p and $P_i^{(p)}(\alpha, c)$ is an appropriate normalization factor (such that all individuals have the same number of contacts, regardless of geography and citizenship, see Appendix for further details). Each contribution consists of an *endogenous* term, accounting for the infectivity due to individuals from the same population, and an *exogeneous* term, accounting for the infectivity due to the other population. The interplay between mobility and social integration is thus encoded in the force of infection for the two populations.

APPENDIX

3.A METHODS

3.A.1 *Human mobility patterns in Turkey*

The mobility of Turkish and Syrian refugees in the country is modeled as follows. We consider two populations of individuals, namely population 1 of size N^1 and population 2 of size N^2, living in a territory consisting of K geographically separated patches, corresponding to the Turkish prefectures. Each patch $k \in \{1,...,K\}$ accounts for N_k^1 and N_k^2 individuals, with $\sum_{k=1}^{K} N_k^1 = N^1$ and $\sum_{k=1}^{K} N_k^2 = N^2$. The absolute number of individuals moving between patches is inferred from available Call Detail Records following an approach similar to one used in previous studies [93, 100] and rescaled in such a way to adequately represent the the current amount of individuals living in Turkey, which consists in 80M Turkish citizens and 3.5M Syrian refugees. Let c_{ik}^1 be the daily number of individuals of population 1 travelling from patch i to patch k, with $\sum_{k=1}^{K} c_{ik}^1 = N_i^1$ and c_{ii}^1 representing individuals who remain in the patch i. Similarly, let c_{ik}^2 be the daily number of individuals of population 2 travelling from path i to path k, with $\sum_{k=1}^{K} c_{ik}^2 = N_i^2$. The number of individuals in any patch i is therefore defined as:

$$P_i(c) = \sum_{k=1}^{K} c_{ki}^1 + \sum_{k=1}^{K} c_{ki}^2.$$

3.A.2 *The measles transmission model*

Measles transmission is considered as driven by a deterministic meta-population transmission model following a Susceptible-Infected-Removed schema [6, 69, 106, 151, 124], but taking into account: i) heterogeneous immunity among Turkish citizens and Syrian refugees; ii) realistic individuals' mobility patterns across different Turkish prefectures, characterizing both the Turkish and the Syrian populations; iii) the effect of different levels of social integration between the two populations. We assume that, in each considered prefecture, individuals of population 1 mix homogeneously among themselves and with a fraction α (with $0 \leq \alpha \leq 1$) of individuals

of population 2. As a consequence, the overall contact rate of each individual of the population 1 in patch i is assumed proportional to:

$$P_i^1(\alpha, c) = \sum_{k=1}^{K} c_{ki}^1 + \alpha \sum_{k=1}^{K} c_{ki}^2. \tag{5}$$

Similarly, the contact rate of individuals of population 2 in patch i is assumed proportional to:

$$P_i^2(\alpha, c) = \alpha \sum_{k=1}^{K} c_{ki}^1 + \sum_{k=1}^{K} c_{ki}^2. \tag{6}$$

It is worth noting that, according to this definition of contacts, $\alpha = 0$ represents the situation of two completely separated populations (therefore mimicking the full segregation of refugees in the country) and $\alpha = 1$ corresponds to fully homogeneous mixing among individuals of the two populations.

We assume that the contact rate of each individual is equal to a certain value σ for all individuals. Basically, this means that the individual contact rate does not depend on population type, mobility, mixing, and geography. According to this assumption, the following equations should be satisfied in any patch i:

$$\sigma = \sigma^{*1} P_i^1(\alpha, c); \qquad \sigma = \sigma^{*2} P_i^2(\alpha, c),$$

for individuals of population 1 and 2, respectively. It is straightforward to see that this occurs by setting:

$$\sigma_i^{*1}(\alpha, c) = \sigma / P_i^1(\alpha, c); \qquad \sigma_i^{*2}(\alpha, c) = \sigma / P_i^2(\alpha, c).$$

The rate of contacts of individuals of population 1 with infected individuals can be therefore defined as:

$$Q_i^1(\alpha, c) = \sigma_i^{*1}(\alpha, c) \left[\sum_{k=1}^{N} c_{ki}^1 \frac{I_k^1}{N_k^1} + \alpha \sum_{k=1}^{K} c_{ki}^2 \frac{I_k^2}{N_k^2} \right],$$

where $\frac{I_k^1}{N_k^1}$ and $\frac{I_k^2}{N_k^2}$ represent the fraction of infected individuals of the two populations in the patch k. Note that the term between square brackets represents the number of infected individuals among $P_i^1(\alpha, c)$. Similarly, the rate of contacts of individuals of population 2 with infected individuals can be defined as:

$$Q_i^2(\alpha, c) = \sigma_i^{*2}(\alpha, c) \left[\alpha \sum_{k=1}^{N} c_{ki}^1 \frac{I_k^1}{N_k^1} + \sum_{k=1}^{K} c_{ki}^2 \frac{I_k^2}{N_k^2} \right],$$

By defining p the probability of infection transmission given a contact, and $\beta = p\sigma$ the measles specific transmission rate, we can assume that susceptible individuals of population 1 in any patch i are exposed to the following force of infection:

$$\lambda_i^1(\alpha, c) = \beta \left[\sum_{k=1}^{K} \frac{c_{ki}^1}{P_i^1(\alpha, c)} \frac{I_k^1}{N_k^1} + \alpha \sum_{k=1}^{K} \frac{c_{ki}^2}{P_i^1(\alpha, c)} \frac{I_k^2}{N_k^2} \right]$$

Similarly, the force of infection for individuals of population 2 can be defined as:

$$\lambda_i^2(\alpha,c)=\beta\left[\alpha\sum_{k=1}^{K}\frac{c_{ki}^1}{P_i^2(\alpha,c)}\frac{I_k^1}{N_k^1}+\sum_{k=1}^{K}\frac{c_{ki}^2}{P_i^2(\alpha,c)}\frac{I_k^2}{N_k^2}\right].$$

Consequently, in our model, the system of ordinary differential equations regulating the epidemic transmission dynamics is the following:

$$\begin{cases}\dot{S}_i^1=-\lambda_i^1(\alpha,c)S_i^1\\ \dot{S}_i^2=-\lambda_i^2(\alpha,c)S_i^2\\ \dot{I}_i^1=\lambda_i^1(\alpha,c)S_i^1-\gamma I_i^1\\ \dot{I}_i^2=\lambda_i^2(\alpha,c)S_i^2-\gamma I_i^2\\ \dot{R}_i^1=\gamma I_i^1\\ \dot{R}_i^2=\gamma I_i^2\end{cases}$$

for $i\in\{1,...,K\}$, where $\gamma^{-1}=14$ days is the exponentially distributed generation time associated with measles infection [6].

It is worth noting that the approach developed and used here to model measles transmission in Turkey is fairly general to be also applied to other infectious diseases and other socio-demographic settings.

3.A.3 *Initial contitions*

We assume different levels of protection against the infection f^1 and f^2 $(0\le f^j\le 1)$ for individuals of the populations 1 and 2, respectively. We also assume that, at time $t=0$, the epidemic is seeded by one single index case randomly chosen in a given patch. If the index case is an individual belonging to population 1 and living in patch i^*, the initial conditions in patch i^* are $S_{i^*}^1(0)=(1-f^1)N_{i^*}^1-1$; $S_{i^*}^2(0)=(1-f^2)N_{i^*}^2$; $I_{i^*}^1(0)=1$; $I_{i^*}^2(0)=0$; $R_{i^*}^1(0)=f^1N_{i^*}^1$; and $R_{i^*}^2(0)=f^2N_{i^*}^2$. In patches $i\neq i^*$, the initial conditions are $S_i^1(0)=(1-f^1)N_i^1$; $S_i^2(0)=(1-f^2)N_i^2$; $I_i^1(0)=0$; $I_i^2(0)=0$; $R_i^1(0)=f^1N_i^1$; and $R_i^2(0)=f^2N_i^2$.

3.A.4 *Reproduction numbers*

Reproduction numbers associated to the epidemic transmission model and defining the transmission potential of measles in Turkey are computed by applying next generation matrix techniques [39, 41, 103]. Specifically, if we define $X=(I_1^1,...,I_K^1,I_1^2,...,I_K^2)$ and $Y=(S_1^1,...,S_K^1,S_1^2,...,S_K^2)$, it is straightforward to observe that, for appropriate choices of coefficients m_{ik}, the model equations for X can be written in the form

$$\dot{X}_i=Y_i\sum_{k=1}^{2K}p\sigma m_{ik}\frac{X_k}{N_k}-\gamma X_i,$$

where m_{ik} are defined as:

$$\begin{aligned}
m_{ik} &= \frac{c^1_{ki}}{P^1_i(\alpha,c)} \text{ for } i \in \{1,...,K\} \text{ and } k \in \{1,...,K\},\\
m_{ik} &= \frac{\alpha c^2_{ki}}{P^1_i(\alpha,c)} \text{ for } i \in \{1,...,K\} \text{ and } k \in \{K+1,...,2K\},\\
m_{ik} &= \frac{\alpha c^1_{ki}}{P^2_i(\alpha,c)} \text{ for } i \in \{K+1,...,2K\} \text{ and } k \in \{1,...,K\},\\
m_{ik} &= \frac{c^2_{ki}}{P^2_i(\alpha,c)} \text{ for } i \in \{K+1,...,2K\} \text{ and } k \in \{K+1,...,2K\}.
\end{aligned}$$

Note that the terms σm_{ik} represent numbers of contacts that individuals in patch i have with individuals of patch k and thus, put in this form, the model resembles a classical age structured SIR model. By denoting with $\mathbf{M}$ the matrix with entries m_{ik}, it follows that:

$$R_0 = p\rho(\sigma\mathbf{M})\gamma^{-1},$$

where $\rho(\sigma\mathbf{M})$ indicates the spectral radius of $\sigma\mathbf{M}$. Since $\mathbf{M}$ is a probability matrix (also termed transition matrix, i.e. all rows sum up to 1) it follows that $\rho(\mathbf{M}) = 1$, and therefore $R_0 = \beta\gamma^{-1}$, resembling the R_0 associated with simple homogeneous mixing SIR models.

The effective reproduction number can be computed in a similar way, but accounting for the susceptibility of infectors, that is by defining $\mathbf{M}$ as the matrix with entries:

$$\begin{aligned}
m^*_{ik} &= m_{ik}(1-f^1)\ (i = 1,...,K;\ k = 1,...,K)\\
m^*_{ik} &= m_{ik}(1-f^2)\ (i = 1,...,K;\ k = K+1,...,2K)\\
m^*_{ik} &= m_{ik}(1-f^1)\ (i = K+1,...,2K;\ k = 1,...,K)\\
m^*_{ik} &= m_{ik}(1-f^2)\ (i = K+1,...,2K\ k = K+1,...,2K).
\end{aligned}$$

3.A.5 *Estimating measles immunity levels among Turkish citizens and Syrian refugees*

Two different immunity levels against measles infection are assumed in the Turkish and the Syrian populations, hereafter denoted by f^1 and f^2 respectively. As measles epidemics have not been recently reported in Turkey, we assume that f^1 reflects the fraction of immunized individuals among recent birth cohorts through 1^{st} and 2^{nd} dose routine vaccination programs. In particular, by assuming a vaccine efficacy $e = 95\%$ [154] and considering the average coverage levels for the 1^{st} and 2^{nd} doses reported by the WHO for the period 2006-2016, $c_1 = 97\%$ and $c_2 = 88\%$ respectively [166], we estimate $1 - f^1$ as:

$$1 - f^1 = 1 - c_1 + c_1(1-e)(1-c_2) + c_1c_2(1-e)^2$$

where $1-c_1$ denotes the fraction of individuals who have never been vaccinated, $c_1(1-e)(1-c_2)$ represents the fraction of individuals who have been vaccinated only with 1^{st} dose but have experienced vaccine failure (occurring in a fraction 1-e of vaccinees), and $c_1c_2(1-e)^2$ defines the fraction of individuals who experienced vaccine failure after 2 dose administrations.

Estimates of f^2 were obtained by inferring the fraction of susceptible individuals among different age classes in the Syrian population, on the basis of data on the measles epidemic reported in Syria during 2017 (> 700 cases) [165], and accounting for the age distribution of Syrian refugees in Turkey. More specifically, we estimate the effective reproductive number (R_e) associated with the 2017 epidemic as $R_e = 1 + r\,T_g$, where $T_g = 14$ days is the measles generation time [6], r is the exponential growth rate in the weekly number of cases reported during 2017 (Fig. 26 A,B). Estimates of R_e are used to derive the fraction of susceptible individuals in Syria at the beginning of 2017 as $S_0 = R_e/R_0$, using three values of R_0: 12, 15 and 18 [69, 90, 124]. Estimates of S_0 are combined with the age distribution of observed cases and the age structure of the considered population (Fig. 26 C,E) to estimate the age specific immunity profile of the Syrian population. Specifically, the fraction of immune individuals in each age group a (Fig. 26 D) is approximated as:

$$imm(a) = 1 - S_0 \frac{cases(a)}{\sum cases} \frac{\sum pop}{pop(a)},$$

where $cases(a)$ denotes the number of cases observed in the age group a, $\sum cases$ denotes the total number of cases of any age, $pop(a)$ is the number of individuals of age a in the population and $\sum pop$ is the overall population size. Finally, the fraction of susceptible Syrian refugees in Turkey was obtained by combining the age specific immunity profile estimated for Syria and the age distribution of Syrian refugees in Turkey [1] (Fig. 26 E). Estimates obtained on the fraction of measles susceptible refugees (i.e. $1-f^2$) are shown in Fig. 26 F.

The obtained results suggest that the effective reproductive number (R_e) of the recent Syrian measles epidemic was 1.32 (95%CI 1.26–1.38), the percentage of susceptible individuals in Syria at the beginning of the 2017 measles epidemic (S0) was 8.92 (95%CI 7.29–10.96); consequently, the percentage of susceptible individuals among Syrian refugees in Turkey was estimated to be 9.87% (95%CI 8.07—12.18).

3.A.6 *Simulating measles epidemic in Turkey*

Simulations of measles epidemics in Turkey were obtained under three different scenarios of $R_0 = 12, 15, 18$, three different scenarios of $1-f^2 = 0.0987, 0.0807, 0.1218$ and eleven illustrative values of α. Explored values of α were selected in such a way to reproduce different proportion of Syrian contacts occurring with Turkish citizens: from 0% to 92%; the former representing the full segregation scenario and the latter repre-

senting homogeneous mixing between Syrian refugees and Turkish citizens. Starting from Eq. 5 and 6 it is easy to see that, for any given value of α, the fraction of contacts that Syrian refugees have with Turkish citizens in patch i is given by $\alpha B_i/(A_i+\alpha B_i)$, where A_i and B_i are the number of Syrian refugees and Turkish citizens in patch i respectively. It follows that the average fraction of contacts that Syrian refugees have with Turkish citizens in the whole study area is given by $\sum_i w_i\, \alpha B_i/(A_i+\alpha B_i)$, where $w_i = A_i/\sum_i A_i$. Similarly, the average fraction of contacts that Turkish citizens have with Syrian refugees in the whole study area is given by $\sum_i \tilde{w}_i\, \alpha A_i/(B_i+\alpha A_i)$, where $\tilde{w}_i = B_i/\sum_i B_i$. Note that for a given value of α the fraction of Syrian contacts with Turkish citizens is generally different from the fraction of Turkish contacts with the Syrians. The value of α resulting in a certain fraction x of contacts of Syrian refugees with Turkish citizens can therefore be computed by solving the equation

$$\sum_i w_i \frac{\alpha B_i}{A_i+\alpha B_i} = x.$$

For each considered scenario, 100 measles epidemics were simulated for one year starting from the September, by seeding each epidemic in one different patch, selected among the 100 prefectures of Turkey with the highest amount of refugees. In our simulations, the force of infection associated with measles transmission is assumed to decrease by a factor $\tilde{r}$ during summer (i.e. between June and September), as a consequence of school closure. Following estimates provided in [124], $\tilde{r}$ was taken equal to 0.33. Finally, all simulations were performed under the assumption of $1-f^1 = 0.038$. However, an illustrative value of $1-f^1 = 0.05$ was also explored for sensitivity analysis.

3.B ADDITTIONAL RESULTS

The developed transmission model takes into account both mobility patterns and different levels of social mixing between Syrian and Turkish population. Here, we summarize the results obtained by simulating a measles epidemic in Turkey for all the considered scenarios associated with different policies for the integration of Syrian refugees. Figure 31 shows the probability of experiencing an outbreak causing at least 20 cumulative infections among the Syrian or the Turkish population. Figure 32 displays the number of prefectures affected by those simulated epidemic that cause at least 20 cumulative infections overall (i.e. in the sum of the two populations). Obtained results show that higher levels of integration are associated with a lower probability of measles re-emergence in Turkey. Additionally, our analysis also show that increasing the social integration of the two population could favor the containment of potential spread of the infection, resulting in a smaller amount of prefectures affected by the measles epidemic.

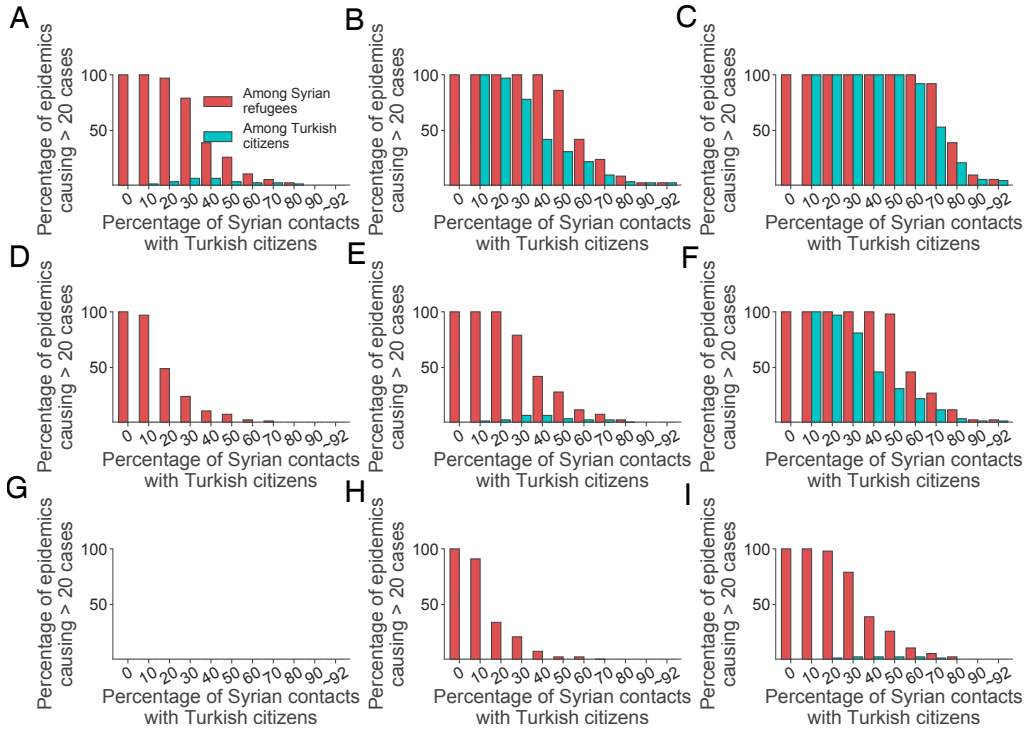

Figure 31: Bars represent the percentage of epidemics causing at least 20 cases among among Syrian refugees (red) and Turkish citizens (blue) for the model projections as a function of the mixing parameter. Three different values of R_0 and susceptibility levels among the Syrian refugees population ($1-f^2$) were considered. **A)** $R_0 = 18$, $1-f^2 = 0.0807$. **B)** as A but for $1-f^2 = 0.0987$. **C)** as A but for $1-f^2 = 0.1218$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

3.B.1 *Sensitivity analysis*

We also investigate the role played by the uncertainty surrounding the immunity level against measles infection assumed in the Turkish population in determining the robustness of obtained results. We therefore simulate all scenarios explored so far under the assumption that 5% of Turkish citizens are susceptible (i.e. $1-f^1 = 0.05$). Obtained results show that, although in this case benefits coming from social integration of refugees in Turkey would result slightly reduced (Fig. 33, 34, 35, S36), the importance of social integration for reducing epidemic risk associated with measles in Turkey is strongly confirmed.

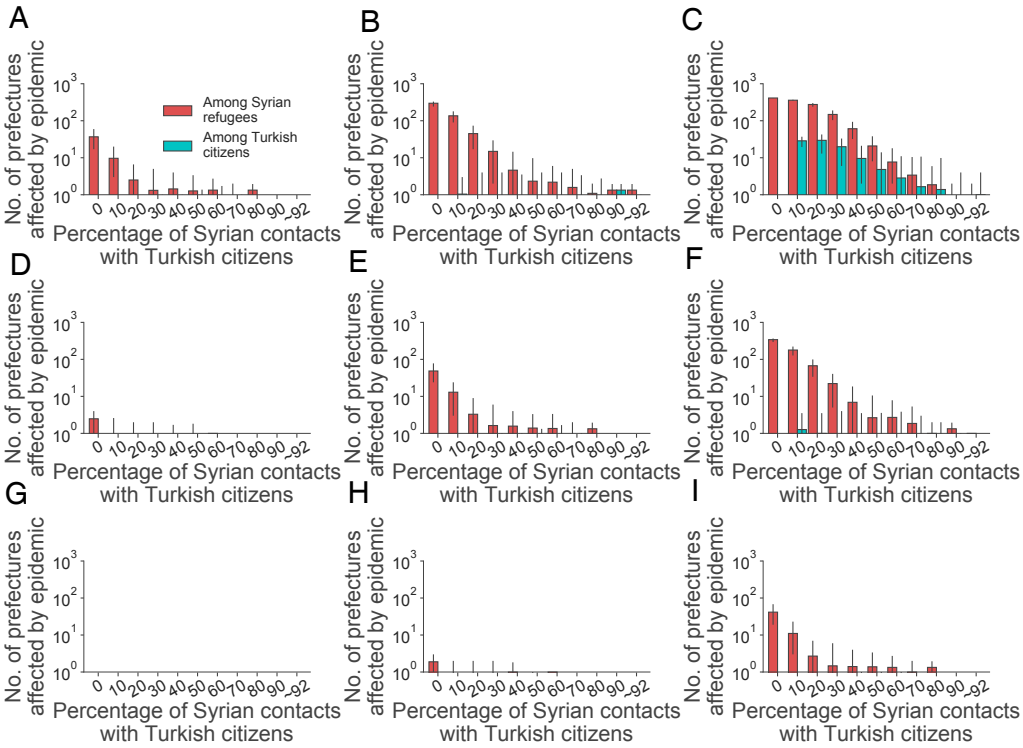

Figure 32: Bars represent the average number of prefectures exceeding 20 cases among Syrian refugees (red) and Turkish citizens (blue); black lines indicate the 95%CI. Three different values of R_0 and susceptibility levels among the Syrian refugees population ($1-f^2$) were considered. **A)** $R_0 = 18$, $1-f^2 = 0.0807$. **B)** as A but for $1-f^2 = 0.0987$. **C)** as A but for $1-f^2 = 0.1218$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

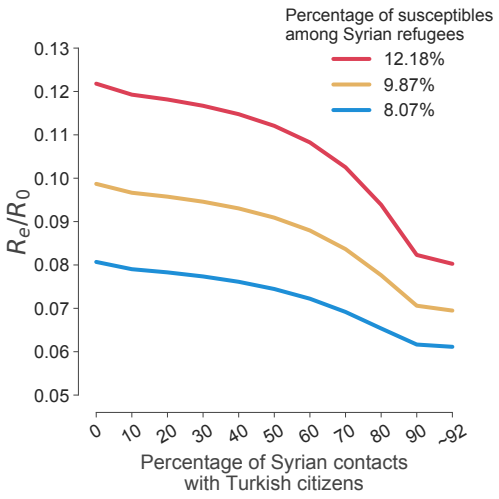

Figure 33: Effective reproduction number for measles spreading according to our model, rescaled by R_0, as a function of the mixing parameter accounting for social integration between Turkish and Refugees as obtained by assuming $1 - f^1 = 0.05$. Colored lines are associated with the estimated levels of susceptibility among Syrian refugees.

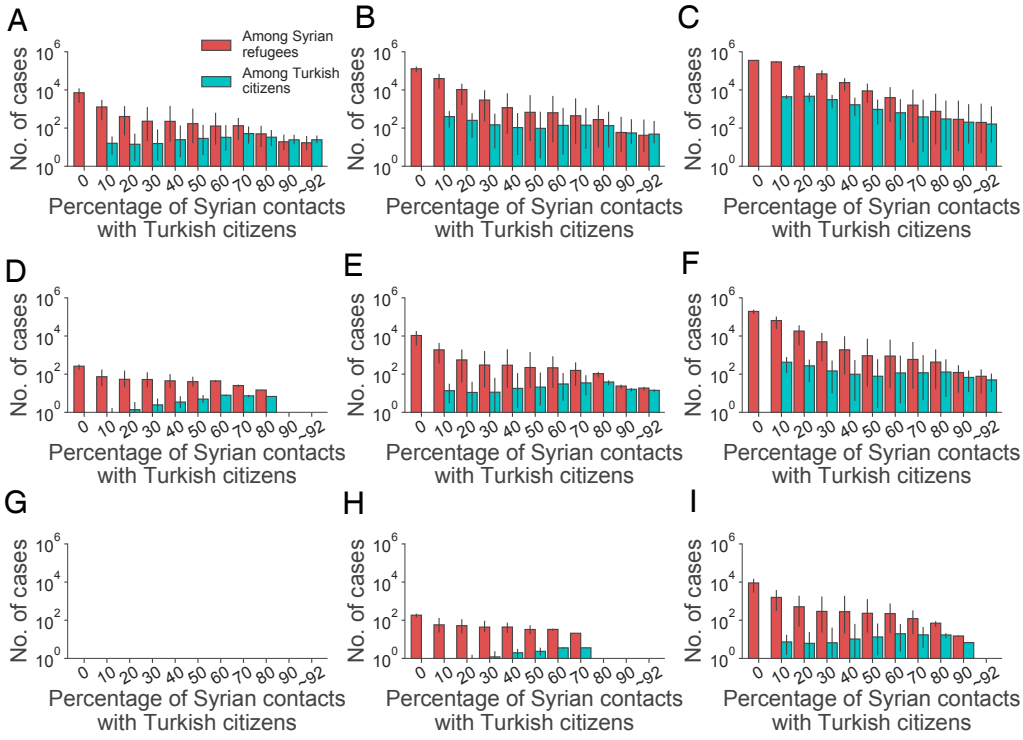

Figure 34: Cumulative infections as obtained by assuming $1 - f^1 = 0.05$ and considering epidemics that exceed 20 cases in the entire population. Bars represent the average number of infections occurring among Syrian refugees (red) and Turkish citizens (blue) for the model projections as a function of the mixing parameter, black lines represent 95%CI. Three different values of R_0 and Susceptibility levels among the Syrian refugees population (S_0^{Ref}) were considered. **A)** $R_0 = 18$, $S_0^{Ref} = 8.07\%$. **B)** as A but for $S_0^{Ref} = 9.87\%$. **C)** as A but for $S_0^{Ref} = 12.15\%$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

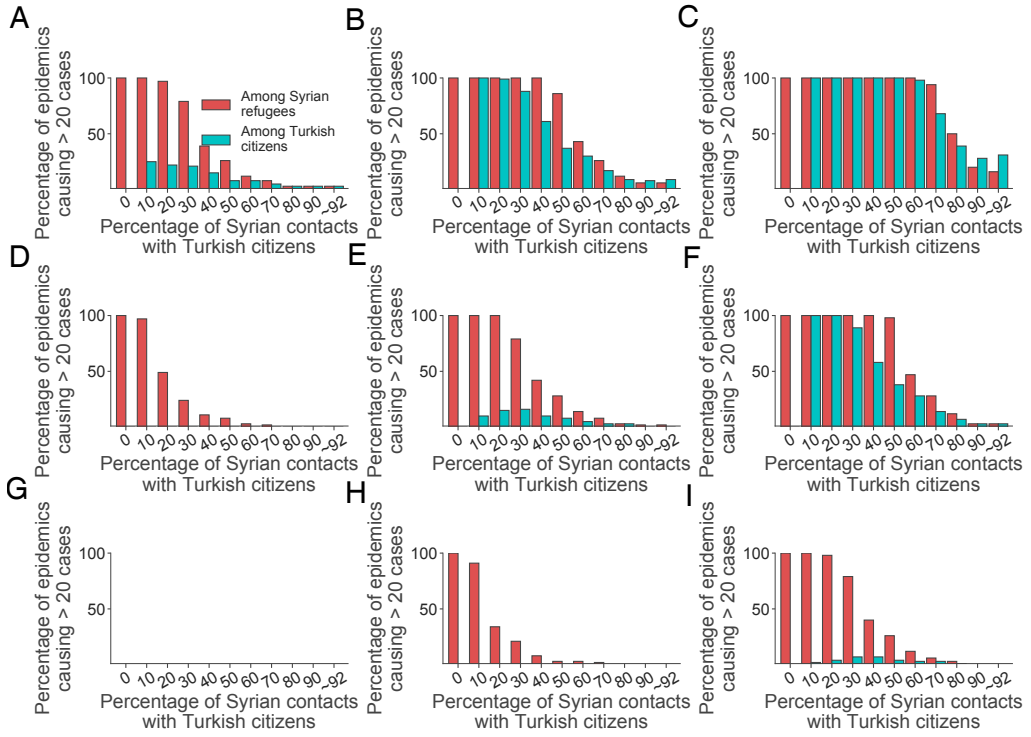

Figure 35: Bars represent the percentage of epidemics causing at least 20 cases among among Syrian refugees (red) and Turkish citizens (blue) for the model projections as a function of the mixing parameter, as obtained by assuming $1-f^1 = 0.05$. Three different values of R_0 and susceptibility levels among the Syrian refugees population $(1-f^2)$ were considered. **A)** $R_0 = 18$, $1-f^2 = 0.0807$. **B)** as A but for $1-f^2 = 0.0987$. **C)** as A but for $1-f^2 = 0.1218$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

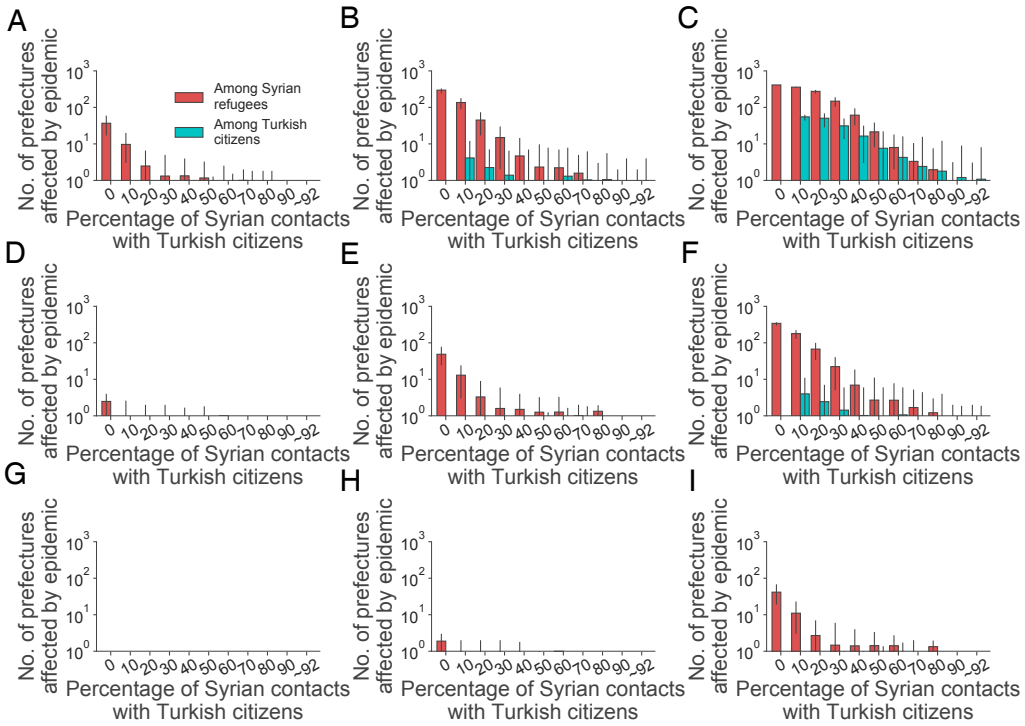

Figure 36: Bars represent the average number of prefectures exceeding 20 cases among Syrian refugees (red) and Turkish citizens (blue) as obtained by assuming $1-f^1 = 0.05$; black lines indicate the 95%CI. Three different values of R_0 and susceptibility levels among the Syrian refugees population ($1-f^2$) were considered. **A)** $R_0 = 18$, $1-f^2 = 0.0807$. **B)** as A but for $1-f^2 = 0.0987$. **C)** as A but for $1-f^2 = 0.1218$. **D)** as A but for $R0 = 15$. **E)** as B but for $R0 = 15$. **F)** as C but for $R0 = 15$. **G)** as A but for $R0 = 12$. **H)** as B but for $R0 = 12$. **I)** as C but for $R0 = 12$.

4

CONCLUSION

The interplay between human behavior and epidemic spreading represents a challenge in describing the dynamics of infectious diseases [98, 57, 59, 155]. As extensively discussed, understanding the role of human behavior, and how deeply it affects the impact and diffusion of an epidemic outbreak, represents a crucial step. In this thesis, we focused on two aspects involving the dynamics of infectious diseases associated with human behaviors.

The first aspect is related to possible behavioral changes that occur as a response to pandemic events. As a matter of fact, a change in people's behavior can be either spontaneous, in order to reduce their risk of infection, or imposed by the government as a necessary response to protect the community [48, 48, 27, 145]. In both cases, risk perception and awareness of the epidemic is a crucial aspect that should be considered to characterize the reactive response to a health threat [159]. On one hand, awareness triggers spontaneous behavioral changes that may break the disease transmission. On the other hand, it promotes the public acceptance of the restrictive measures imposed by policy makers, making them more effective [2, 44, 4]. Awareness is often invoked in models of infectious disease dynamics, it is not clear which are the main drivers responsible for the increased awareness in the population. Actually, awareness is usually coupled with the disease prevalence, in fact neglecting possible mechanism among biological and social contagion processes. Thus, understanding these types of mechanism would allow to conceive models that explicitly account for the effects of population awareness into the progression of infectious diseases in a population.

The second problem we investigated is associated with the possibility of measles re-emergence in Turkey caused by social dynamics, i.e. the segregation of Syrian refugees, in the hosting population. Since few years, the number of migrants has been increasing worldwide with a trend constantly growing. Migrants are nowadays perceived more and more often as a threat for the countries health system and economy [3]. This negative perception triggers mechanisms of social segregation of migrant and refugees and may cause a cascade of adverse consequences detrimental to social stability, and potentially dangerous for the public health of the country [31]. As an emblematic case study, we considered Turkey, currently hosting 3.5 million of Syrian refugees and facing a humanitarian emergency of unprecedented levels. In fact, one of the most serious concerns of the Turkish population is about the possibility of dis-

ease outbreaks, such as measles, because of the poor immunization coverage in Syria due to recent civil war in the last years.

In chapter 2 we developed a transmission model to detect the major drivers of awareness spreading associated with the emergence of an epidemic threat. The considered model is based on a contagion process that considers two routes of transmission, namely world of mouth social contagion and mass media communication. Contagion processes are extensively used in the literature to model infectious disease dynamic but recently were used also to model other types of phenomena such as contagion of non-healthy behaviors, emotions, and ideas [30, 76, 16]. We took advantage of novel data sources such as the number of Wikipedia page views over time and the volume of news article collected in Google News per day about a specific disease. Specifically, Wikipedia page usage was employed as a proxy of the awareness in the population, while Google News data was assumed to mirror the volume of media response to the considered epidemic.
We analyzed four different epidemic threats (Ebola, Zika, Meningitis epidemics and seasonal influenza) in two countries, i.e. the United States and Italy, to asses the contribution of the different information drivers in the awareness building of infectious diseases.

The analysis carried out suggests that the proposed model is able to disentangle the contribution of news influence and word of mouth social contagion in driving the Wikipedia accesses in the case of different public health threats. Furthermore, it provides better explanatory and forecasting power than alternative models considering only one of the driving mechanisms. Most importantly, the model allows the measurement of parameters defining the contagion process such as the fraction of aware people and the relative contributions of the two different contagion processes.

In chapter 3 we build a data driven computational transmission model to evaluate the likelihood of experiencing a measles outbreak in Turkey. In particular, the model is informed with mobility patterns inferred from mobile phone data and considers different levels of immunity among Syrian refugees and Turkish citizens. Immunity levels in the refugees' population are estimated from the recent measles outbreak occurred in Syria during 2017. The model we propose is capable to evaluate different scenarios of integration/segregation between the two populations that are associated with the possibility of measles re-emergence in the country. Results strongly suggest that improving integration of Syrian refugees can mitigate and contain the diffusion of measles in Turkey with respect to the two different populations. Thanks to the use of mobile phone data, we are able to estimate the spatial distribution of Syrian refugees in Turkey and to characterize the prefecture responsible for the onset of possible widespread epidemics. Accordingly, the developed analysis suggests that

the areas most at risk for seeding of future measles epidemics may be those with a high percentage of Syrian refugees in the population, as a higher proportion of susceptible individuals in the population can sustain the initial spread of the epidemic at the local level. These results can be used to guide targeting immunization campaigns in specific areas characterized by a large number of refugees with respect to the Turkish population.

As pointed out, the digital revolution has opened the door for new methodologies to reveal the hidden structures and behaviors of our society [131]. This work represents a step in this direction. We have used data on individual usage of the internet to gain information about the major drivers of awareness during health emergence over time. Moreover, mobile phone data were adopted to asses refugees behavioral patterns recently emerging as a consequence of the political crisis in Syria. Accordingly, collecting real time data represent a great opportunity to identify and explore novel elements characterizing the dynamics [87]. Feeding mathematical models with these new data streams is a natural way to exploit this information. However, data on individual behaviors are difficult to collect and obtain, thus more effort is needed to to allow both replications of experiments and comparative studies in the community of infectious diseases dynamics [74].

As a final remark, a general limitation, that might globally affect the investigated dynamics in our works, is the difficulty in modeling the heterogeneity in the individual contacts. The large diversity both in the number and in the intensity of individual contacts might be responsible for additionally hidden patterns that we do not consider in our study. In particular, information about an epidemic may be conveyed by opinion leaders to the less active segment of the population [86]. Identifying opinion leaders can be useful also to target cost-effective awareness campaigns [159]. In addition, potential differences in measles transmissibility in Syrian refugees communities and in Turkish citizens, induced by different numbers of overall contacts, can affect the predictions on the possibility of the disease re-emergence. Thus, the understanding of complex phenomena would benefit from the availability of new sources of data, characterizing the individual behaviors in a deeper level, such as information on the opinion leaders in a population or age-specific social contact pattern in migrants communities.

Concluding, models benefit from the different contributions coming from a broad spectrum of research fields. This rich interdisciplinary knowledge contributes in turn to the performances and reliability of the predictions. The dialogue between scientists and policy makers is ultimately needed to shape the right and effective strategies in the fight of infectious diseases.

[1] Distribution by age and gender of registered Syrian refugees. Technical report, Republic Of Turkey - Minister of the Interior. Accessed: 2017-08-01.

[2] Sharon Alane Abramowitz, Kristen E McLean, Sarah Lindley McKune, Kevin Louis Bardosh, Mosoka Fallah, Josephine Monger, Kodjo Tehoungue, and Patricia A Omidian. Community-centered responses to ebola in urban liberia: the view from below. *PLoS neglected tropical diseases*, 9(4):e0003706, 2015.

[3] Ibrahim Abubakar, Robert W Aldridge, Delan Devakumar, Miriam Orcutt, Rachel Burns, Mauricio L Barreto, Poonam Dhavan, Fouad M Fouad, Nora Groce, Yan Guo, et al. The ucl–lancet commission on migration and health: the health of a world on the move. *The Lancet*, 392(10164):2606–2654, 2018.

[4] Philip Baba Adongo, Philip Teg-Nefaah Tabong, Emmanuel Asampong, Joana Ansong, Magda Robalo, and Richard M Adanu. Beyond knowledge and awareness: addressing misconceptions in ghana's preparation towards an outbreak of ebola virus disease. *PloS one*, 11(2):e0149627, 2016.

[5] Marco Ajelli, Piero Poletti, Alessia Melegaro, and Stefano Merler. The role of different social contexts in shaping influenza transmission during the 2009 pandemic. *Sci Rep*, 4:7218, 2014.

[6] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[7] Andrea Apolloni, Chiara Poletto, José J Ramasco, Pablo Jensen, and Vittoria Colizza. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11(1):3, 2014.

[8] Ronan F Arthur, Emily S Gurley, Henrik Salje, Laura SP Bloomfield, and James H Jones. Contact structure, mobility, environmental impact and behaviour: the importance of social forces to infectious disease dynamics and disease ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1719):20160454, 2017.

[9] Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic. *PloS one*, 6(1):e16591, 2011.

[10] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA*, 106(51):21484–21489, 2009.

[11] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The

global epidemic and mobility computational model. *Journal of computational science*, 1(3):132–145, 2010.

[12] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics*, 7(7):581, 2011.

[13] Binnur Balkan, Elif Tok, Huzeyfe Torun, and Semih Tumen. Immigration, housing rents, and residential segregation: Evidence from syrian refugees in turkey. *Institute of Labor Economics*, 2015.

[14] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 2018.

[15] Chris T Bauch and Alison P Galvani. Social factors in epidemiology. *Science*, 342(6154):47–49, 2013.

[16] Luís MA Bettencourt, Ariel Cintrón-Arias, David I Kaiser, and Carlos Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364:513–536, 2006.

[17] Sue Binder, Alexandra M Levitt, Jeffrey J Sacks, and James M Hughes. Emerging infectious diseases: public health issues for the 21st century. *Science*, 284(5418):1311–1313, 1999.

[18] Brad K Blitz, Alessio d'Angelo, Eleonore Kofman, and N. Montagna. Health challenges in refugee reception: Dateline europe 2016. *Int J Environ Res Public Health*, 14(12):1484, 2017.

[19] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ data science*, 4(1):10, 2015.

[20] David E Bloom, Michael Kuhn, and Klaus Prettner. Health and economic growth. 2018.

[21] Joshua Blumenstock. Don't forget people in the use of big data for development, 2018.

[22] Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

[23] Martin CJ Bootsma and Neil M Ferguson. The effect of public health measures on the 1918 influenza pandemic in us cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.

[24] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[25] Cristian Candia, C Jara-Figueroa, Carlos Rodriguez-Sickert, Albert-László Barabási, and César A Hidalgo. The universal decay of collective memory and attention. *Nature Human Behaviour*, page 1, 2018.

[26] Stephen Castles, Hein De Haas, and Mark J Miller. *The age of migration: International population movements in the modern world*. Macmillan International Higher Education, 2013.

[27] Simon Cauchemez, Neil M Ferguson, Claude Wachtel, Anders Tegnell, Guillaume Saour, Ben Duncan, and Angus Nicoll. Closure of schools during an influenza pandemic. *The Lancet infectious diseases*, 9(8):473–481, 2009.

[28] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[29] Pauline H Cheong, Rosalind Edwards, Harry Goulbourne, and John Solomos. Immigration, social cohesion and social capital: A critical review. *Critical Social Policy*, 27(1):24–49, 2007.

[30] Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[31] Jocalyn Clark and Richard Horton. Opening up to migration and health. *The Lancet*, 2018.

[32] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA*, 103(7):2015–2020, 2006.

[33] BJ Cowling, LM Ho, and GM Leung. Effectiveness of control measures during the sars epidemic in beijing: a comparison of the r t curve and the epidemic curve. *Epidemiology & Infection*, 136(4):562–566, 2008.

[34] Hippolyte d'Albis, Ekrame Boubtane, and Dramane Coulibaly. Macroeconomic evidence suggests that asylum seekers are not a "burden" for western european countries. *Science Advances*, 4(6):eaaq0883, 2018.

[35] Manlio De Domenico, Clara Granell, Mason A Porter, and Alex Arenas. The physics of spreading processes in multilayer networks. *Nature Physics*, 12(10):901, 2016.

[36] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013.

[37] Sara Del Valle, Herbert Hethcote, James M Hyman, and Carlos Castillo-Chavez. Effects of behavioral changes in a smallpox attack model. *Mathematical biosciences*, 195(2):228–251, 2005.

[38] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem. Dynamic population mapping using mobile phone data. *Proc Natl Acad Sci USA*, 111(45):15888–15893, 2014.

[39] O Diekmann, JAP Heesterbeek, and MG Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, page rsif20090386, 2009.

[40] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.

[41] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990.

[42] Peter Sheridan Dodds and Duncan J Watts. Universal behavior in a generalized model of contagion. *Physical review letters*, 92(21):218701, 2004.

[43] Cem Ergun and Hüseyin Gül. Urban regeneration and social segregation: The case of istanbul. *Toplum ve Demokrasi Dergisi*, 5(11-12), 2014.

[44] Shannon M Fast, Sumiko Mekaru, John S Brownstein, Timothy A Postlethwaite, and Natasha Markuzon. The role of social mobilization in controlling ebola virus in lofa county, liberia. *PLoS currents*, 7, 2015.

[45] Eli P Fenichel, Carlos Castillo-Chavez, M Graziano Ceddia, Gerardo Chowell, Paula A Gonzalez Parra, Graham J Hickling, Garth Holloway, Richard Horan, Benjamin Morin, Charles Perrings, et al. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311, 2011.

[46] Neil Ferguson. Capturing human behaviour. *Nature*, 446(7137):733, 2007.

[47] Neil M Ferguson, Derek AT Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke.

Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209, 2005.

[48] Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448, 2006.

[49] Paul EM Fine. Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302, 1993.

[50] International Organization for Migration. World migration report 2018. 2018.

[51] Wikimedia Foundation. Pagecount raw.

[52] Wikimedia Foundation. Pagecount raw.

[53] Wikimedia Foundation. Pageviews per language.

[54] Laura Fumanelli, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput Biol*, 8(9):e1002673, 2012.

[55] Isaac Chun-Hai Fung, Zion Tsz Ho Tse, Chi-Ngai Cheung, Adriana S Miu, and King-Wa Fu. Ebola and the social media. 2014.

[56] S Funk, E Gilad, and VAA Jansen. Endemic disease, awareness, and local behavioural response. *Journal of theoretical biology*, 264(2):501–509, 2010.

[57] Sebastian Funk, Shweta Bansal, Chris T Bauch, Ken TD Eames, W John Edmunds, Alison P Galvani, and Petra Klepac. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*, 10:21–25, 2015.

[58] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.

[59] Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.

[60] Alison P Galvani and Robert M May. Epidemiology: dimensions of superspreading. *Nature*, 438(7066):293, 2005.

[61] Paul A Gastañaduy, Jeremy Budd, Nicholas Fisher, Susan B Redd, Jackie Fletcher, Julie Miller, Dwight J McFadden III, Jennifer Rota, Paul A Rota, Carole Hickman, et al. A measles outbreak in an underimmunized amish community in ohio. *New England Journal of Medicine*, 375(14):1343–1354, 2016.

[62] Paul A Gastañaduy, Sebastian Funk, Prabasaj Paul, Lilith Tatham, Nicholas Fisher, Jeremy Budd, Brian Fowler, Sietske de Fijter, Mary DiOrio, Gregory S Wallace, et al. Impact of public health responses during a measles outbreak in an amish community in ohio: modelling the dynamics of transmission. *American journal of epidemiology*, 2018.

[63] Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, Sara Y Del Valle, and Reid Priedhorsky. Global disease monitoring and forecasting with wikipedia. *PLoS computational biology*, 10(11):e1003892, 2014.

[64] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[65] Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M Elizabeth Halloran, and Alessandro Vespignani. Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLoS currents*, 6, 2014.

[66] J Gómez-Gardeñes, D Soriano-Paños, and A Arenas. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nature Physics*, 14(4):391, 2018.

[67] Marta C González, Cesar H Hidalgo, and Albert-László Barabási. Understanding individual human mobility. *Nature*, 453:779–782, 2008.

[68] Google. Google News.

[69] Rebecca F Grais, MJ Ferrari, C Dubray, ON Bjørnstad, BT Grenfell, A Djibo, F Fermon, and Philippe J Guerin. Estimating transmission intensity for a measles epidemic in niamey, niger: lessons for intervention. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(9):867–873, 2006.

[70] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical review letters*, 111(12):128701, 2013.

[71] Simon Gregson, Geoffrey P Garnett, Constance A Nyamukapa, Timothy B Hallett, James JC Lewis, Peter R Mason, Stephen K Chandiwana, and Roy M Anderson. Hiv decline associated with behavior change in eastern zimbabwe. *Science*, 311(5761):664–666, 2006.

[72] Sunetra Gupta, Roy M Anderson, and Robert M May. Networks of sexual contacts: implications for the pattern of spread of hiv. *AIDS (London, England)*, 3(12):807–817, 1989.

[73] Sally Hargreaves. Concerns in turkey about infections from refugees. *Lancet Infect Dis*, 16(7):782–783, 2016.

[74] Hans Heesterbeek, Roy M Anderson, Viggo Andreasen, Shweta Bansal, Daniela De Angelis, Chris Dye, Ken TD Eames, W John Edmunds, Simon DW Frost, Sebastian Funk, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339, 2015.

[75] Herbert W Hethcote, James A Yorke, and Annett Nold. Gonorrhea modeling: a comparison of control methods. *Mathematical Biosciences*, 58(1):93–109, 1982.

[76] Alison L Hill, David G Rand, Martin A Nowak, and Nicholas A Christakis. Emotions as infectious diseases in a large social network: the sisa model. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1701):3827–3835, 2010.

[77] James M Hyman and E Ann Stanley. Using mathematical models to understand the aids epidemic. *Mathematical Biosciences*, 90(1-2):415–473, 1988.

[78] Felana Angella Ihantamalala, Vincent Herbreteau, Feno MJ Rakotoarimanana, Jean Marius Rakotondramanga, Simon Cauchemez, Bienvenue Rahoilijaona, Gwenaëlle Pennober, Caroline O Buckee, Christophe Rogier, CJE Metcalf, et al. Estimating sources and sinks of malaria parasites in madagascar. *Nature communications*, 9(1):3897, 2018.

[79] Niall PAS Johnson and Juergen Mueller. Updating the accounts: global mortality of the 1918-1920" spanish" influenza pandemic. *Bulletin of the History of Medicine*, pages 105–115, 2002.

[80] James Holland Jones and Marcel Salathe. Early assessment of anxiety and behavioral response to novel swine-origin influenza a (h1n1). *PLoS one*, 4(12):e8032, 2009.

[81] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990, 2008.

[82] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.

[83] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[84] Jaro Kotalik. Preparing for an influenza pandemic: ethical issues. *Bioethics*, 19(4):422–431, 2005.

[85] AJ Kucharski and Christian Althaus. The role of superspreading in middle east respiratory syndrome coronavirus (mers-cov) transmission. *Euro surveillance*, 20(25):pii–21167, 2015.

[86] Paul Felix Lazarsfeld, Bernard Berelson, and Hazel Gaudet. The people's choice. 1944.

[87] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721, 2009.

[88] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.

[89] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[90] J Lessler, WJ Moss, SA Lowther, and DAT Cummings. Maintaining high rates of measles immunization in africa. *Epidemiology & Infection*, 139(7):1039–1049, 2011.

[91] Sheng Li, Chao Ma, Lixin Hao, Qiru Su, Zhijie An, Fubao Ma, Shuyun Xie, Aiqiang Xu, Yanyang Zhang, Zhengrong Ding, et al. Demographic transition and the dynamics of measles in six provinces in china: A modeling study. *PLoS medicine*, 14(4):e1002255, 2017.

[92] Qiuyan Liao, Benjamin J Cowling, Wendy WT Lam, Diane MW Ng, and Richard Fielding. Anxiety, worry and cognitive risk estimate in relation to protective behaviors during the 2009 influenza a/h1n1 pandemic in hong kong: ten cross-sectional surveys. *BMC infectious diseases*, 14(1):169, 2014.

[93] A Lima, Manlio De Domenico, V Pejovic, and M Musolesi. Disease containment strategies based on mobility and information dissemination. *Scientific Reports*, 5:10650, 2015.

[94] James O Lloyd-Smith, Dylan George, Kim M Pepin, Virginia E Pitzer, Juliet RC Pulliam, Andrew P Dobson, Peter J Hudson, and Bryan T Grenfell. Epidemic dynamics at the human-animal interface. *science*, 326(5958):1362–1367, 2009.

[95] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355, 2005.

[96] Xin Lu, Linus Bengtsson, and Peter Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proc Natl Acad Sci USA*, 109(29):1576–1581, 2012.

[97] Xin Lu, David J Wrathall, Pål Roe Sundsoy, Mohammed Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimer Qureshi, Andrew Tatem, Geoffrey Canright, Kenth Engo-Monsen, and Linus Bengtsson. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in bangladesh. *Global Environmental Change*, 38:1–7, 2016.

[98] Piero Manfredi and Alberto D'Onofrio. *Modeling the interplay between human behavior and the spread of infectious diseases*. Springer Science & Business Media, 2013.

[99] Emanuele Massaro, Alexander Ganin, Nicola Perra, Igor Linkov, and Alessandro Vespignani. Resilience management during large-scale epidemic outbreaks. *Scientific reports*, 8(1):1859, 2018.

[100] Joan T Matamalas, Manlio De Domenico, and Alex Arenas. Assessing reliable human mobility patterns from higher order memory in mobile communications. *Journal of The Royal Society Interface*, 13(121):20160203, 2016.

[101] David J McIver and John S Brownstein. Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time. *PLoS computational biology*, 10(4):e1003581, 2014.

[102] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[103] Alessia Melegaro, Emanuele Del Fava, Piero Poletti, Stefano Merler, Constance Nyamukapa, John Williams, Simon Gregson, and Piero Manfredi. Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *PLoS One*, 12(1):e0170459, 2017.

[104] Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports*, 1:62, 2011.

[105] Stefano Merler and Marco Ajelli. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc Biol Sci*, 277(1681):557–565, Feb 2010.

[106] Stefano Merler and Marco Ajelli. Deciphering the relative weights of demographic transition and vaccination in the decrease of measles incidence in Italy. *Proc Biol Sci*, 281(1777), Feb 2014.

[107] Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis L Chao, Ira M Longini Jr, M Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*, 15(2):204–211, 2015.

[108] Stefano Merler, Marco Ajelli, Andrea Pugliese, and Neil M Ferguson. Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. *PLoS Comput Biol*, 7(9):e1002205, Sep 2011.

[109] David M Morens, Gregory K Folkers, and Anthony S Fauci. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996):242, 2004.

[110] Stephen S Morse. Factors in the emergence of infectious diseases. In *Plagues and politics*, pages 8–26. Springer, 2001.

[111] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008.

[112] Kenrad E Nelson, David D Celentano, Sakol Eiumtrakol, Donald R Hoover, Chris Beyrer, Somboon Suprasert, Surinda Kuntolbutra, and Chirasak Khamboonruang. Changes in sexual behavior and a decline in hiv infection among young men in thailand. *New England Journal of Medicine*, 335(5):297–303, 1996.

[113] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.

[114] OECD. Making integration work: Refugees and others in need of protection. Technical report, Organisation for Economic Co-operation and Development, 2016.

[115] Abhishek Pandey, Katherine E Atkins, Jan Medlock, Natasha Wenzel, Jeffrey P Townsend, James E Childs, Tolbert G Nyenswah, Martial L Ndeffo-Mbah, and Alison P Galvani. Strategies for containing ebola in west africa. *Science*, 346(6212):991–995, 2014.

[116] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[117] K David Patterson and Gerald F Pyle. The geography and mortality of the 1918 influenza pandemic. *Bulletin of the History of Medicine*, 65(1):4–21, 1991.

[118] Umberto Pellecchia, Rosa Crestani, Tom Decroo, Rafael Van den Bergh, and Yasmine Al-Kourdi. Social consequences of ebola containment measures in liberia. *PLoS One*, 10(12):e0143036, 2015.

[119] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PloS one*, 6(8):e23084, 2011.

[120] Robert W Pinner, Steven M Teutsch, Lone Simonsen, Laura A Klug, Judith M Graber, Matthew J Clarke, and Ruth L Berkelman. Trends in infectious diseases mortality in the united states. *Jama*, 275(3):189–193, 1996.

[121] Piero Poletti, Marco Ajelli, and Stefano Merler. The effect of risk perception on the 2009 h1n1 pandemic influenza dynamics. *PloS one*, 6(2):e16460, 2011.

[122] Piero Poletti, Marco Ajelli, and Stefano Merler. Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic. *Mathematical biosciences*, 238(2):80–89, 2012.

[123] Piero Poletti, Bruno Caprile, Marco Ajelli, Andrea Pugliese, and Stefano Merler. Spontaneous behavioural changes in response to epidemics. *Journal of theoretical biology*, 260(1):31–40, 2009.

[124] Piero Poletti, Stefano Parlamento, Tafarraa Fayyisaa, Rattaa Feyyiss, Marta Lusiani, Ademe Tsegaye, Giulia Segafredo, Giovanni Putoto, Fabio Manenti, and Stefano Merler. The hidden burden of measles in ethiopia: how distance to hospital shapes the disease mortality rate. *BMC medicine*, 16(1):177, 2018.

[125] Chiara Poletto, Pierre-Yves Boëlle, and Vittoria Colizza. Risk of mers importation and onward transmission: a systematic review and analysis of cases reported to who. *BMC infectious diseases*, 16(1):448, 2016.

[126] Bernd Rechel, Philipa Mladovsky, David Ingleby, Johan P Mackenbach, and Martin McKee. Migration and health in an increasingly diverse europe. *The Lancet*, 381(9873):1235–1245, 2013.

[127] M Zia Sadique, W John Edmunds, Richard D Smith, William Jan Meerding, Onno De Zwart, Johannes Brug, and Philippe Beutels. Precautionary behavior in response to perceived threat of pandemic influenza. *Emerging infectious diseases*, 13(9):1307, 2007.

[128] Faryad Darabi Sahneh, Fahmida N Chowdhury, and Caterina M Scoglio. On the existence of a threshold for preventive behavioral responses to suppress epidemic spreading. *Scientific reports*, 2:632, 2012.

[129] Albert Ali Salah, Alex Pentland, Bruno Lepri, Emmanuel Letouzé, Patrick Vinck, Yves-Alexandre de Montjoye, Xiaowen Dong, and Özge Dağdelen. Data for refugees: The d4r challenge on mobility of syrian refugees in turkey. *arXiv preprint arXiv:1807.00523*, 2018.

[130] Marcel Salathé and Sebastian Bonhoeffer. The effect of opinion clustering on disease outbreaks. *Journal of The Royal Society Interface*, 5(29):1505–1508, 2008.

[131] Matthew J Salganik. *Bit by bit: social research in the digital age*. Princeton University Press, 2017.

[132] Henrik Salje, Justin Lessler, Kishor Kumar Paul, Andrew S Azman, M Waliur Rahman, Mahmudur Rahman, Derek Cummings, Emily S Gurley, and Simon Cauchemez. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. *Proceedings of the National Academy of Sciences*, 113(47):13420–13425, 2016.

[133] Mauricio Santillana, André T Nguyen, Mark Dredze, Michael J Paul, Elaine O Nsoesie, and John S Brownstein. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, 11(10):e1004513, 2015.

[134] Mauricio Santillana, D Wendong Zhang, Benjamin M Althouse, and John W Ayers. What can digital disease detection learn from (an external revision to) google flu trends? *American journal of preventive medicine*, 47(3):341–347, 2014.

[135] Samuel V Scarpino. Don't close the gates. *Nature Physics*, page 1, 2018.

[136] National Science, Technology Council Pandemic Prediction, Forecasting Science, and Technology Working Group. Towards Epidemic Prediction: Federal Efforts And Opportunities In Outbreak Modeling., 2016.

[137] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96, 2012.

[138] Emily Simons, Matthew Ferrari, John Fricks, Kathleen Wannemuehler, Abhijeet Anand, Anthony Burton, and Peter Strebel. Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data. *The Lancet*, 379(9832):2173–2178, 2012.

[139] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. Why we read wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1591–1600. International World Wide Web Conferences Steering Committee, 2017.

[140] Larry A Sjaastad. The costs and returns of human migration. *Journal of Political Economy*, 70(5, Part 2):80–93, 1962.

[141] Richard D Smith. Responding to global infectious disease outbreaks: lessons from sars on the role of risk perception, communication and management. *Social science & medicine*, 63(12):3113–3123, 2006.

[142] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327:1018–1021, 2010.

[143] Brian G Southwell, Suzanne Dolina, Karla Jimenez-Magdaleno, Linda B Squiers, and Bridget J Kelly. Zika virus–related news coverage and online behavior, united states, guatemala, and brazil. *Emerging infectious diseases*, 22(7):1320, 2016.

[144] David Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika van der Linde. Bayesian measures of model complexity and fit. *Quality control and applied statistics*, 48(4):431–432, 2003.

[145] Gillian K SteelFisher, Robert J Blendon, Mark M Bekheit, and Keri Lubell. The public's response to the 2009 h1n1 influenza pandemic. *New England Journal of Medicine*, 362(22):e65, 2010.

[146] Gillian K SteelFisher, Robert J Blendon, and Narayani Lasala-Blanco. Ebola in the united states—public reactions and implications. *New England Journal of Medicine*, 373(9):789–791, 2015.

[147] Istituto superiore di sanità. Meningite epidemia mediatica, 2017.

[148] Saki Takahashi, C Jessica E Metcalf, Matthew J Ferrari, Andrew J Tatem, and Justin Lessler. The geography of measles vaccination in the african great lakes region. *Nature communications*, 8:15585, 2017.

[149] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Chrsitian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the use of human mobility proxies for modeling epidemics. *Plos Computational Biology*, 10(7):e1003716, 2009.

[150] Sherry Towers, Shehzad Afzal, Gilbert Bernal, Nadya Bliss, Shala Brown, Baltazar Espinoza, Jasmine Jackson, Julia Judson-Garcia, Maryam Khan, Michael Lin, et al. Mass media and the contagion of fear: the case of ebola in america. *PloS one*, 10(6):e0129179, 2015.

[151] Filippo Trentini, Piero Poletti, Stefano Merler, and Alessia Melegaro. Measles immunity gaps and the progress towards elimination: a multi-country modelling analysis. *Lancet Infect Dis*, 17(10):1089–1097, Aug 2017.

[152] UNHCR. 3RP 2017 ANNUAL REPORT Regional Refugee & Resilience Plan 2017-2018. Technical report, United Nations High Commissioner for Refugees, 2018.

[153] Department of Economic United Nations and & Population Division Social Affairs. The international migration report (highlights). 2016.

[154] Amra Uzicanin and Laura Zimmerman. Field effectiveness of live attenuated measles-containing vaccines: a review of published literature. *The Journal of infectious diseases*, 204(suppl_1):S133–S149, 2011.

[155] Frederik Verelst, Lander Willem, and Philippe Beutels. Behavioural change models for infectious disease transmission: a systematic review (2010–2015). *Journal of The Royal Society Interface*, 13(125):20160820, 2016.

[156] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature physics*, 8(1):32, 2012.

[157] Cécile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, et al. The rapidd ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.

[158] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[159] Melanie A Wakefield, Barbara Loken, and Robert C Hornik. Use of mass media campaigns to change health behaviour. *The Lancet*, 376(9748):1261–1271, 2010.

[160] Haidong Wang, Mohsen Naghavi, Christine Allen, Ryan M Barber, Zulfiqar A Bhutta, Austin Carter, Daniel C Casey, Fiona J Charlson, Alan Zian Chen,

Matthew M Coates, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *The lancet*, 388(10053):1459–1544, 2016.

[161] Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch. Coupled disease–behavior dynamics on complex networks: A review. *Physics of life reviews*, 15:1–29, 2015.

[162] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[163] Amy Wesolowski, Taimur Qureshi, Maciej F Boni, Pål Roe Sundsøy, Michael A Johansson, Syed Basit Rasheed, Kenth Engø-Monsen, and Caroline O Buckee. Impact of human mobility on the emergence of dengue epidemics in pakistan. *Proceedings of the National Academy of Sciences*, 112(38):11887–11892, 2015.

[164] W.H.O. WHO Measles Syria, EWARS Weekly Bulletin Week No. 50. Technical report, World Health Organization, 2017.

[165] W.H.O. Global Measles and Rubella Update April 2018, 2018.

[166] W.H.O. Immunization, Vaccines and Biologicals, Data, statistics and graphics, 2018.

[167] W.H.O. WHO Polio Report, Syria cVDPV2 outbreak Situation Report No. 33. Technical report, World Health Organization, 2018.

[168] Maike Winters, Mohamed F Jalloh, Paul Sengeh, Mohammad B Jalloh, Lansana Conteh, Rebecca Bunnell, Wenshu Li, Zangin Zeebari, and Helena Nordenstedt. Risk communication and ebola-specific knowledge and behavior during 2014–2015 outbreak, sierra leone. *Emerging infectious diseases*, 24(2):336, 2018.

[169] Gary Wong, Wenjun Liu, Yingxia Liu, Boping Zhou, Yuhai Bi, and George F Gao. Mers, sars, and ebola: the role of super-spreaders in infectious disease. *Cell host & microbe*, 18(4):398–401, 2015.

[170] Joseph T Wu, Steven Riley, Christophe Fraser, and Gabriel M Leung. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS medicine*, 3(9):e361, 2006.

[171] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore Y Piontti, Natalie E Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, Margaret Bray, M Elizabeth Halloran, Ira M Longini, Jr, and Alessandro Vespignani. Spread of Zika virus in the Americas. *Proc Natl Acad Sci USA*, 114(22):E4334–E4343, Apr 2017.

[172] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore y Piontti, Natalie E Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, et al. Spread of zika virus in the americas. *Proceedings of the National Academy of Sciences*, page 201620161, 2017.

[173] Cathy Zimmerman, Ligia Kiss, and Mazeda Hossain. Migration and health: A framework for 21st century policy-making. *PLoS Med*, 8(5):e1001034, 2011.