Risk factors Linked to Influenza-like Illness as Identified from the Mexican Participatory Surveillance System "Reporta"

Christopher R. Stephens
C3-Centro de Ciencias de la Complejidad and ICN, Universidad Nacional Autonoma de Mexico, CDMX 04510
stephens@nucleares.unam.mx

Rocio Rodríguez-Ramírez
Intellego Business Intelligence, Mexico CDMX, Mexico

Victor Mireles
Facultad de Ciencias, Universidad Nacional Autónoma de Mexico, CDMX 04510

Sergio Hernández-López
C3 and Facultad de Ciencias, Universidad Nacional Autónoma de Mexico, CDMX 04510

Concepción García-Aguirre
C3 and Facultad de Ciencias, Universidad Nacional Autónoma de Mexico, CDMX 04510

Juan Arturo Herrerra-Ortiz
C3-Centro de Ciencias de la Complejidad, Universidad Nacional Autonoma de Mexico, CDMX 04510

Natalia B. Mantilla-Beniers
C3 and Facultad de Ciencias, Universidad Nacional Autónoma de Mexico, CDMX 04510
nmantilla@ciencias.unam.mx

ABSTRACT

Internet-based monitoring of influenza-like illnesses (ILI) has become more common since its beginnings over a decade ago, both through estimates based on the number of searches for influenza-related terms (e.g., Google flu trends), or by means of participatory surveillance systems. The latter, often seen as ways of engaging people in matters of scientific and public health importance, gather a wealth of potentially valuable epidemiological information complementary to that obtained through the established disease surveillance networks and also usually absent from search-based web algorithms.

We present a statistical analysis of the data from the Mexican monitoring website "Reporta" by which the risk factors linked to reporting of ILI symptoms as outcome among its participants are determined, and interpret these results based on current knowledge of the factors that influence transmission of infection resulting in disease. Besides standard factors associated with enhanced susceptibility to infection some novel behavioral factors linked to high risk were: (i) use of public transport; (ii) frequent contact with animals, and (iii) use of non-standard interventions, such as homeopathy. While close contact with large groups of people in public transportation is generally assumed to be important in disease spread, frequent contact with animals is not. Our results are consistent with previous observations that animals may serve as mobile fomites and hence increase the propensity to develop disease. We conclude that analysis of rich information sets from Internet-based systems may suggest novel ideas on disease spread that are worth following up with field research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
DH '17, July 02-05, 2017, London, United Kingdom
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5249-9/17/07…$15.00
http://dx.doi.org/10.1145/3079452.3079471

ACM Reference format:
C.R. Stephens, R. Rodríguez-Ramírez, V. Mireles, S. Hernández-López, C. García-Aguirre, J.A. Herrera-Ortiz, N.B. Mantilla-Berniers. 2017. SIG Proceedings Paper in word Format. In *Proceedings of DH'17,* July 02-05, 2017, London, United Kingdom
© 2017 Association for Computing Machinery. 8 pages.
DOI: http://dx.doi.org/10.1145/3079452.3079471

KEYWORDS

Participatory surveillance system; influenza; risk factors

1 INTRODUCTION

Seasonal epidemics of influenza are estimated to result in three to five million cases of severe illness each year, and between 250,000 to 500,000 deaths [1]. Influenza viruses also have huge potential for emergent pandemics and are important from a basic research point of view, all of which makes them of fundamental public health interest. A reduction in the disease burden caused by influenza requires an understanding of the risk factors associated with transmission and symptom presentation for a given population, while taking interventions into account. The profiles of individuals at high risk of acquiring the infection and developing disease can help to decide where, when and whom to apply an intervention.

In classical epidemiological modeling a common step is to neglect the heterogeneities (in contact, susceptibility, etc.) present in the

process. At the same time, a fundamental question is: When are these heterogeneities of critical importance and therefore must be taken into account? We will consider here the issue of how heterogeneity can impact disease dynamics by embedding it in the conceptual framework of risk analysis and studying which factors determine the risk of contracting the disease for an individual. Because the heterogeneities of disease propagation may be present in a large number of factors, the question arises of how to obtain data associated with them.

Traditionally, the chief sources of epidemiological data have been public health authorities. Many countries have well-developed sentinel systems, whereby the dynamics of an epidemic is tracked by the number of visits to physicians of patients diagnosed with influenza-like illness (ILI) and, occasionally, confirmatory laboratory tests. However, the extent to which the number of doctor visits with ILI diagnosis is a good representation of underlying ILI incidence in the population is rarely known [2]. A variety of cultural, demographic and economic factors may influence whether a person consults a doctor when having ILI symptoms. On the other hand, an advantage of medical records is that they usually include patient information together with a diagnosis, and can be used to study risk factors. Other potential sources for such data have often been clinical studies, where the effect of one, or a small number, of risk factors is studied [3-5], though these are generally costly in both human and monetary resources.

With the advent of the Internet, other sources of data have become available for tracking influenza epidemics [6-13]. A well known one is Google-flu trends [14], whereby frequencies of Internet searches for terms related to influenza are taken as a proxy for ILI incidence. These data have been shown to track sentinel data quite well [10, 14-18]. However, their correlation can decrease abruptly [19, 20], and they are far removed from obtaining data that links ILI incidence to individual traits, thus making an analysis of risk factors unfeasible.

On the other hand, an alternative source of information that has become more popular recently is that of web-based “crowd-sourced" or participatory systems [6, 13, 21, 22]. These have mostly been employed, thus far, to monitor ILI activity through users who voluntarily sign up and then report weekly on whether or not they exhibit ILI symptoms. The participants also input socio-demographic and lifestyle information upon registration. Unlike proxy-based systems akin to Google-flu, participatory surveillance systems (PSS) count the frequency of ILI symptoms reported by the participants. Although self-assessed, this tally is done at the individual level and, by linking it to the background information of each person, it can further our understanding of risk factors in relation to disease spread.

Naturally, the potential utility of any of these different methods for giving epidemiological alerts or contributing to the understanding of risk factors depends on the geographical coverage and sheer numbers of volunteers involved, as well as any intrinsic underlying biases in the population being sampled. On the other hand, all data sources will be biased. In the case of sentinel systems, one bias is that they sample those who are economically able to go to a doctor and live close enough to him. Google-flu-like systems have a bias that comes from measuring interest in flu-like subjects as opposed to ILI incidence itself, and are restricted to Internet users. This last remark is also true for crowd-sourced data. Moreover, since the latter are based on voluntary participation, one may ask whether volunteers are a good representation of the general population, or whether they get sick more/less often, or over-represent a certain sector of the population, thus introducing other potential sources of bias. However, the analyses performed thus far show that voluntary-based crowd-sourced data adequately track the beginning, peak and decay of ILI epidemics [21-24], and when complemented with other sources, their structure allows us to try and correct this bias by taking into account the socio-demographic factors that characterize our sample population. Moreover, PSS provide a way of incorporating changes in healthcare-seeking behaviour that then improve our estimates of disease incidence and severity [25]. We will therefore make the assumption that they are a useful information source for determining and understanding risk factors for influenza spread.

Since our goal is to discover the traits associated to different degrees of risk of ILI, we must consider what causes the emergence of new cases in a population. Two main factors are necessary: coming into contact with the influenza virus, and developing the disease. All else being equal, the first factor depends on the number of contacts one has. The second factor depends chiefly on characteristics of the individual, such as immune response (itself dependent on a wide variety of factors, such as history of infection and immunization, nutrition, genetic background, presence of chronic conditions, etc.) and personal hygiene. This last group of factors ultimately defines the tendency to acquire the infection and develop disease, which we term - individual susceptibility. Thus, we will: (a) measure the associations between the background data and ILI presentation found in the Mexican website “Reporta”, (b) provide a “taxonomy" in which background data are classified as influencing the contact network or reflecting differences in susceptibility. Lastly, we will (c) interpret risk levels in terms of our proposed taxonomy. Our interpretation is then substantiated with the results of a refined assessment of risk and leads to hypotheses that may be tested in the lab or in the field.

2 Methods

2.1 Study Population

The population for this study was participants of the crowd-sourced flu-monitoring system “Reporta” (web site http://reporta.c3.org.mx/) which opened in Mexico just after the H1N1 sanitary emergency there in 2009. Such systems rely on direct, voluntary, weekly reporting of symptoms to produce an estimate of the prevalence of ILI among participants [21]. As such, the sample population is not necessarily representative of the full Mexican population, and so our conclusions apply only to the sample population considered and any extrapolation to a wider population is subject to considerations of sample representativeness. In the specific case of Reporta, its genesis within the National Autonomous University of Mexico resulted in a study population skewed towards age groups and occupations particularly associated with a university as can be seen in Figure 1. However, the profile is not dissimilar to that of participants of other crowd-sourced systems [22].

Figure 1. Age profile of Reporta participants in contrast to that of the general population in Mexico (INEGI).

2.2 Ethical Considerations

This research was approved by the Ethics Committee of the National Institute of Respiratory Diseases "Ismael Cossio Villegas" of the National Health Ministry of Mexico (INER/CEI/052/2013). Written, informed consent was obtained for the analysis of anonymous data. In order to register, participants must tick a box by which they give their consent for taking part in a scientific study on influenza epidemiology. The consent form details how their data will be used and states that their information will remain anonymous and confidential. It also explains how they may end their participation.

2.3 Sample Size, Study Period and Outcome Measure

We took into account data from N = 4,873 Reporta participants, a sample made up by those who registered, completed their background questionnaire and filled in one or more weekly surveys between the launching of the system in May 2009 and September 2011. This sample was chosen in order to focus on the epidemiology of the H1N1 strain, which was presumed to dominate influenza epidemiology at the time. The principal outcome measure was the occurrence of ILI in any participant over the study period. Our classification of a set of reported symptoms as indicative of an ILI follows the definition used by the Mexican Health Ministry (Secretaría de Salud or SSA), which requires that the patient present, at least, a fever (more than 38C) and either a cough, or throat pain. We focus on the class, C, of respondents who reported ILI symptoms at any time during the interval from the start of their participation to September 2011 ($P(C) = N_C/N = 20.09\%$), where $N_C = 979$. Just as with other Internet-based monitoring systems one can compare PSS-based ILI incidence with official statistics [21-24] in order to determine the degree of correlation. In Figure 2 we show a comparison of ILI incidence among participants of Reporta against hospital discharge data. As can be observed, the tendencies are quite similar.

2.4 REPORTA: Questionnaire and variables

Participant data consist of their answers to: (i) a background questionnaire that is filled in immediately after registration, and

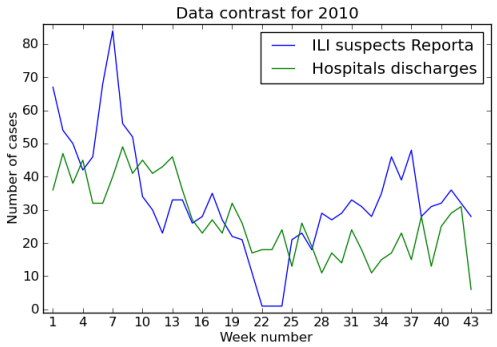

Figure 2. Contrast of influenza-like illness among participants of Reporta and pneumonia discharges from hospitals. The time series portrayed in this graph show the weekly numbers of participants of Reporta with ILI symptoms who live in the Federal District of Mexico (D.F.) and those of pneumonia patients discharged from hospitals of the Health Ministry of D.F.

(ii) weekly symptoms questionnaires. User ID numbers identify questionnaires that correspond to a single participant. The background questionnaire of Reporta consists of 21 questions covering a range of socio-demographic and lifestyle traits. The background questionnaire asks what is the approximate number of colds that participants have each year, with three possible answers in a drop-down menu: less than two, between two and five, more than five colds per year. Participants are also classified as belonging to a risk group when they are older than 65 or under 2, or if they reported having any chronic disease. Another question asks what the respondent does when falling ill (to which possible replies are "Visit the GP" or "Take what the pharmacist recommends", among other options in a list). It is also enquired whether the participant has frequent contact with a variety of animals from a list, or with "other" animals that do not appear specifically in it.

An important issue for crowd-sourced systems is the frequency of participation, which we measure by considering the fraction of weekly questionnaires that each participant fills in, counted from the date in which the participant first registered in the system. Thus, a 50% participation rate could correspond to someone who has been in the system one month and filled in two reports, or someone who has been in the system two years and filled in 52 reports. We note that well over 50% of participants completed more than 30% of their weekly questionnaires.

2.5 Statistical Analysis

We will consider the identification of risk factors, X_i, for the class, C = presence of ILI. As a classification problem we wish to

determine the relation between C and X_i through $P(C|X_i)$. Hence, a natural statistical diagnostic is the binomial test

$$\varepsilon(C \mid X_i) = N_{Xi} (P(C \mid X_i) - P(C)) / (N_{Xi}P(C)(1 - P(C)))^{1/2}$$

which determines the statistical significance of the relation between the class, C = presence of ILI, and the risk factors X_i, where N_{Xi} is the number of participants with the risk factor X_i. We will also consider the odds ratio $OR = [P(C|X_i)/P(\underline{C}|X_i)]/[P(C|\underline{X_i})/P(\underline{C}|\underline{X_i})]$, where $\underline{C}$ is the complement of C and $\underline{X_i}$ the complement of X_i.

In this paper we concentrate on a univariate analysis. We also performed bivariate analysis on the data, which generally support the univariate results by showing that the factors identified with high (and low) risk combine in producing profiles of the same risk category [27] (results not shown). To determine the confidence intervals for the binomial test we use the Wilson score interval [28]. Note that in contrasting $P(C|X_i)$ to $P(C)$ we have chosen the weaker null hypothesis associated with the class C, because C contains the elements CX_i, which are also in X_i. The stronger null hypothesis $P(C|\underline{X_i})$, where $\underline{X_i}$ represents those values that are the complement of X_i, would exclude the part of C made up by participants who also are in X_i. As $N_{CXi} << N_C$ the difference between using C and $C\underline{X_i}$ is negligible. We have checked explicitly with both null hypotheses and the set of significant factors remains the same. The identification of those factors X_i that are most correlated with ILI incidence allows us to establish a risk profile for any person for whom some or all of the same data is available. Consequently, new participants now receive a preliminary risk assessment that is based on our analyses of the initial part of the data upon registration.

3 RESULTS

In Tables 1 and 2 we list the most significant risk factors: in terms of increased (Table 1), or of reduced (Table 2) risk of presenting ILI symptoms. The traits listed are those for which $\varepsilon(C|X_i)$ corresponds to the 95% confidence level using the Wilson intervals. Note that all potential risk factors were evaluated, but only those that exhibited statistically significant correlation with ILI incidence (28 out of 171, 16.37%) are displayed. As mentioned, we also calculated ε for all two-variable combinations, but did not find any conclusions that were qualitatively different from those of the univariate analysis.

Trait	P(C|X)	**Upper 95% CI**	**Lower 5% CI**	ε	OR
More than five colds per year	33.20%	25.60%	15.52%	5.08	2.06
Belongs to a risk group	25.51%	22.37%	17.99%	4.85	1.54
2-5 colds per year	24.52%	22.02%	18.29%	4.66	1.53
Chronic respiratory disease	29.21%	24.27%	16.47%	4.57	1.73
"Other" chronic disease*	29.05%	28.40%	13.74%	4.41	1.71
Treats health issues with herbal/home remedies	24.44%	22.49%	17.88%	3.69	1.40
Transport: bus	23.12%	21.88%	18.42%	3.43	1.38
Transport: foot	24.63%	22.87%	17.57%	3.35	1.38
Contact with "other" domestic or farm animals	25.78%	23.78%	16.85%	3.21	1.44
Age range (25,30]	24.65%	23.05%	17.43%	3.17	1.37
Contact with birds	24.73%	23.15%	17.34%	3.12	1.38
Uses homeopathy	25.17%	23.57%	17.01%	3.04	1.40
Does not use medical services	27.60%	25.49%	15.59%	2.96	1.56
Female	22.23%	21.63%	18.63%	2.8	1.36
Marital status: single	21.95%	21.80%	18.48%	2.2	1.24
Contact with cats	22.43%	22.38%	18.51%	2.09	1.21

(* excludes diabetes, respiratory and heart conditions)

Table 1. Most significant risk factors for ILI symptoms

In Table 1 we can observe a natural grouping of the most significant risk factors: (i) a group of high susceptibility factors, shown in yellow, corresponding to those who have a higher disease burden in general, either in terms of colds or in terms of chronic diseases, and those who reported belonging to a risk group; (ii) a group of factors, in blue, corresponding to the type of treatment that participants seek, indicating in this case a tendency to self-medicate or to seek non-standard alternatives; (iii) a group of factors, in magenta, corresponding to frequent contact with animals; (iv) a group of factors, in green, corresponding to the type and number of transport that the

participant uses; and (v) a group of demographic factors, in white: age, gender and marital status. The same color code is used in other tables.

Trait	P(C|X)	**Upper 95% CI**	**Lower 5% CI**	ε	OR
Does not use herbal or home remedies	18..02%	22.14%	18.19%	-2.05	0.71
Does not have contact with dogs	18.08%	21.41%	18.83%	-2.23	0.81
Occupation: Industrial	12.72%	21.91%	18.38%	-2.42	0.57
Age range (60,65]	12.12%	26.68%	14.80%	-2.55	0.54
Male	17.64%	21.87%	18.42%	-2.78	0.76
Does not belong to a risk group	18.15%	21.43%	18.81%	-2.90	0.65
Does not use bus	17.87	21.61%	18.65%	-2.93	0.72
Does not have contact with animals	16.54%	22.44%	17.93%	-3.09	0.73
Less than two colds per year	16.48%	21.64%	18.63%	-4.71	0.60

Table 2. Most significant factors for low frequency presentation of ILI symptoms.

In Table 2 we see that participants who rarely present ILI symptoms belong, as might be expected, to classes that are generally complementary to those of the high-risk group. The reduced risk age group 60-65 is, in fact, consistent with the pattern observed elsewhere [29,30], as is the higher attack rate of H1N1 among younger age groups [29, 31]. It has been explained by preexisting immunity to a related strain among the older age groups, which would have been exposed at the time in which it circulated. Since our interest is to understand what these results can tell us from an epidemiological point of view, we proceed to make some concrete inferences and hypotheses that further work could validate.

3 DISCUSSION

A non-infected individual can only become infected through contact with influenza virions, something that depends on the individual's contact network. For the contact to result in infection, the virus must then reach the tissues or organs by which it can enter the host. Thus, contagion is also linked to personal and environmental hygiene. Lastly, the actual outcome of an infection, whether it will result in the manifestation of certain symptoms, is often linked with individual susceptibility. Susceptibility can refer to ease in acquiring infection or in developing disease; here, we will concentrate on the latter, which is reflected in the symptoms recorded in our data. In light of the information available in Reporta, we propose a taxonomy of the factors associated to risk of presenting ILI symptoms by linking these traits, when possible, to: (a) contact network, and (b) susceptibility of an individual.

The variables deemed to serve as proxies for distinct risks of **contact** are: age (since it is linked to a social context, e.g., attending school, retirement), occupation (someone who works at home is less likely to have a large network of contacts than someone who works in customer service), type and number of transports used, household size, and being in frequent contact with various animals (these may be thought of as fomites or vehicles for disease which effectively extend the contact network).

Those variables used as proxies for differential **susceptibility** are: age, occupation, number of colds per year (as recorded by the participant in the initial questionnaire), whether vaccinated against seasonal flu and why, number of hours devoted to physical exercise per week, preexisting chronic conditions, belonging to a risk group, and whether the respondent is a smoker. We also could potentially classify frequent contact with animals as a susceptibility factor, given its potential association to allergic responses.

Markers for differential susceptibility are reported directly by the participants, who indicate how many colds per year they tend to present (we classify those participants who claim to have more than five colds per year as highly susceptible), whether they belong to a risk group, and whether they have any chronic conditions. All of these traits were linked to high ε values, as shown in Table 1.

The results there also suggest that frequent contact with animals might carry enhanced risk of infection. We hypothesize that pets may effectively extend the contact network, serving as mobile fomites. Handling or petting may deposit pathogens on the animal, which then pass on, on the fur, to other members of the household. Personal communication with veterinarians provides two pieces of evidence in support of this hypothesis: First, collateral results in experimental studies report findings of causative agents (Staphylococcus) of human respiratory diseases in animal fur (Escorcia-Martínez, M., personal communication). Second, research on the source of a viral infection in animals has demonstrated that arthropods may mechanically carry the virus and have the potential to transmit it to other hosts [32, 33]. If this hypothesis were correct, one would expect that, all else being equal, having contact with a greater number of animal types should lead to an enhanced risk of presenting ILI symptoms. We tested this hypothesis with the results seen in Table 3, which clearly confirm the hypothesis.

Contact with number of animal types	P(C|X)	Upper 95%	Lower 95%	ε
0	16.54%	18.73%	14.56%	0
1	19.77%	18.13%	15.06%	4.12
2	21.96%	19.01%	14.33%	4.54
3 or more	27.12%	20.32%	13.34%	5.94

Table 3: Risk of presenting ILI symptoms associated to the number of animal types with which participants are in frequent contact. This risk is calculated in contrast to the frequency of ILI symptoms among participants who do not have frequent contact with animals of any kind. Because of the different point of contrast of this analysis, the results below are not part of those reported in Table 1. The columns of this table use the same data as those of Table 1.

Similarly, one might imagine that people who use a greater number of transport types are more likely to have a larger contact network than those who only use one. Our data support this idea, as risk values do indeed increase monotonically with the number of transport types used as seen in Table 1. Interestingly, contrasting results have appeared in the literature related to the correlation between transport and incidence of ILI. In [34] for instance, it was shown that there was no significant correlation between ILI and public transport usage for crowd-sourced data from the Netherlands, Belgium and Portugal. On the other hand, in [35] it was shown that there was a correlation between bus/tram usage and incidence of acute respiratory infection. Notably, in [35], however, the correlation was found between bus/tram usage within 5 days of symptom onset as opposed to the typical frequency of bus/tram usage.

4 CONCLUSIONS

A number of studies have shown that crowd-sourced data can mirror the start, peak and decline of ILI incidence in the general population, and have exhibited their potential for estimating underreporting and improving various parameter estimates [21 - 25]. They have thus been validated as a method of epidemiological data collection that provides novel, useful information.

The present study exploits a novel data set that contains both epidemiological data and a wealth of sociodemographic and lifestyle details to understand what risk factors differentiate individuals in terms of their probability of presenting ILI symptoms. The approach reveals interesting correlations between ILI presence and potential risk factors that may then be interpreted in the light of epidemiological and biological knowledge.

Some of our findings, such as the association between the number of transport types typically used and increasing risk of presenting ILI symptoms, would tend to confirm that high rates of mixing generally increase the likelihood of contracting infections [36]. On the other hand, the similar link of said risk to frequent contact with animals brings attention to a lesser known factor, and raises the question of what mechanism lies behind this connection. We hypothesize that contact with animals may be a way in which the effective contact network is extended and infection risk enhanced, as a result of animals acting as mobile fomites that have close contact with household members. Alternatively, animal hair or feathers may cause irritation of the airways that then are more prone to pathogen colonization. We found that the number of animal types with which an individual is in frequent contact is positively associated to the likelihood of presenting ILI symptoms. Without being conclusive, this would favor the hypothesis that animals extend the contact network.

Either way, it is clear that the integrated, wide-ranging information provided by crowd-sourced data may yield new links and insights into old phenomena. In turn, these new hypotheses may be investigated further. The observations documented in this paper may also be complemented and contrasted against other studies on similar data. In the same way in which extensive data sets of disease incidence have revealed the relative importance of various demographic factors [37, 38] and the role of particular pathogen life history traits on disease dynamics [39], crowd-sourcing can yield large, rich data banks that already are considerably easier to process and analyze than printed records. We believe that data-mining is worth pursuing beyond mere descriptions, and have proposed a preliminary taxonomy of demographic and lifestyle traits that may serve as an intuitive interpretive framework linked to classical notions of epidemiology.

ACKNOWLEDGMENTS

This work exists only because of the generous participation of people who have registered in the website and donated some of their time and data on a weekly basis. We are grateful to them. We are also indebted to Dr. Jorge G. Morales Velazquez, Director of Health Information at the Health Ministry of the Federal District, for kindly lending us the hospital discharge records used in the comparison against data from Reporta. Reporta is based in the C3 - Centro de Ciencias de la Complejidad of the National Autonomous University of Mexico (UNAM). This work was supported by the Instituto de Ciencia y Tecnologia del Distrito Federal (ICyT-DF) of Mexico (grants No. ICyTDF/60/2010 and ICyTDF/343/2010), PAPIIT grant IN113414 and CONACyT Fronteras grant 2015-2-1093.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

REFERENCES

[1] World Health Organization. Influenza (Seasonal). Fact sheet 211, June 2009.

[2] C.C. van den Wijngaard, W. van Pelt, N.J. Nagelkerke, M. Kretzschmar, and M.P. Koopmans. Evaluation of syndromic surveillance in The Netherlands: Its added value and recommendations for implementation. Euro Surveillance, 16(9), 2011.

[3] KL Nichol, A. Lind, KL Margolis, M. Murdoch, R. McFadden, M. Hauge, S. Magnan, and M. Drake. The effectiveness of vaccination against influenza in healthy, working

adults. New England Journal of Medicine, 333(14):889-893, OCT 5 1995 1995.

[4] TME Govaert, CTMCN Thijs, N. Masurel, MJW Sprenger, GJ Dinant, and JA Knottnerus. The efficacy of influenza vaccination in elderly individuals - a randomized double-blind placebo-controlled trial. JAMA-Journal of the American Medical Association, 272(21):1661-1665, DEC 7 1994 1994.

[5] PA Gross, AW Hermogenes, HS Sacks, J. Lau, and RA Levandowski. The efficacy of influenza vaccine in elderly persons - A metaanalysis and review of the literature. Annals of Internal Medicine, 123(7):518-527, OCT 1 1995 1995.

[6] Y. Vandendijck, C. Faes, and N. Hens. Eight years of the Great Influenza Survey to monitor influenza-like illness in Flanders. PLoS ONE, 8(5), 2013.

[7] Q. Yuan, E.O. Nsoesie, B. Lv, G. Peng, R. Chunara, and J.S. Brownstein. Monitoring influenza epidemics in china with search query from Baidu. PLoS ONE, 8(5), (2013).

[8] C.D. Corley, D.J. Cook, A.R. Mikler, and K.P. Singh. Using web and social media for influenza surveillance. Advances in Experimental Medicine and Biology, 680: 559-564, 2010. (1996).

[9] L.C. Mado. Promed-mail: An early warning system for emerging diseases. Clinical Infectious Diseases, 39(2): 227-232, (2004).

[10] G. Eysenbach. Infodemiology: tracking u-related searches on the web for syndromic surveillance. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, pages 244-248, 2006.

[11] U. Bilge, S. Bozkurt, B.O. Yolcular, and D. Ozel. Can social web help to detect influenza related illnesses in Turkey? volume 174, pages 100{104, 2012.

[12] A. Hulth, G. Rydevik, and A. Linde. Web queries as a source for syndromic surveillance. PloS ONE, 4(2), 2009.

[13] J.S. Brownstein, C.C. Freifeld, B.Y. Reis, and K.D. Mandl. Surveillance sans frontieres: Internet-based emerging infectious disease intelligence and the healthmap project. PLoS Medicine, 5(7):1019-1024, 2008.

[14] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. Nature, 457(7232):1012-4, 2009.

[15] J.R. Ortiz, H. Zhou, D.K. Shay, K.M. Neuzil, A.L. Fowlkes, and C.H. Goss. Monitoring influenza activity in the United States: A comparison of traditional surveillance systems with google flu trends. PLoS ONE, 6(4), 2011.

[16] M.T. Malik, A. Gumel, L.H. Thompson, T. Strome, and S.M. Mahmud. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. Canadian Journal of Public Health, 102(4):294-297, 2011.

[17] K. Wada, H. Ohta, and Y. Aizawa. Correlation of "google flu trends" with sentinel surveillance data for influenza in 2009 in Japan. Open Public Health Journal, 4:17{20, 2011.

[18] A. Valdivia, J. Lopez-Alcalde, M. Vicente, M. Pichiule, M. Ruiz, and M. Ordobas. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks - results for 2009-10. Euro surveillance: Bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 15(29), 2010.

[19] Declan Butler. When Google got flu wrong. Nature, 494(7436): 155-156, February 2013.

[20] Samantha Cook, Corrie Conrad, Ashley L. Fowlkes, and Matthew H. Mohebbi. Assessing google flu trends performance in the united states during the 2009 influenza virus a (h1n1) pandemic. PloS ONE, 6(8):e23610, 08 2011.

[21] SP van Noort, M Muehlen, H Rebelo de Andrade, C Koppeschaar, JM Lima Lourenco, and Gomes MG. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. Euro Surveillance, 12(7), 2007.

[22] Natasha L. Tilston, Ken T. D. Eames, Daniela Paolotti, Toby Ealden, and W. John Edmunds. Internet-based surveillance of influenza-like-illness in the uk during the 2009 h1n1 influenza pandemic. BMC Public Health, 10:650, OCT 27 2010 2010.

[23] Richard L. Marquet, Aad I. M. Bartelds, Sander P. van Noort, Carl E. Koppeschaar, John Paget, Francois G. Schellevis, and Jouke van der Zee. Internet-based monitoring of influenza-like illness (ILI) in the general population of The Netherlands during the 2003-2004 influenza season. BMC Public Health, 6:242, OCT 4 2006 2006.

[24] I. H. M. Friesema, C. E. Koppeschaar, G. A. Donker, F. Dijkstra, S. P. van Noort, R. Smallenburg, W. van der Hoek, and M. A. B. van der Sande. Internet-based monitoring of influenza-like illness in the general population: Experience of five influenza seasons in the netherlands. VACCINE, 27(45):6353-6357, OCT 23 2009.

[25] Ellen Brooks-Pollock, Natasha Tilston, W. John Edmunds, and Ken T. D. Eames. Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England. BMC Infectious Diseases, 11:68, MAR 16 2011 2011.

[26] D. Hand, H. Mannila, and P. Smyth. Principles of Data Mining. Cambridge, MA: MIT Press, 2001.

[27] Rocio Rodriguez Ramirez. Analisis del Sistema Ciudadano de Monitoreo de Enfermedades Respiratorias “Reporta” con mineria de datos. Bachelor's thesis, Facultad de Ciencias, Universidad Nacional Autonoma de Mexico, August 2012.

[28] S.A. Wallis. Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. Journal of Quantitative Linguistics, 20(3):178-208, 2013.

[29] Elizabeth Miller, Katja Hoschler, Pia Hardelid, Elaine Stanford, Nick Andrews, and Maria Zambon. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. The Lancet, 375(9720):1100-1108, April 2010.

[30] H.A. Kelly, G.N. Mercer, J.E. Fielding, G.K. Dowse, K. Glass, D. Carcione, K.A. Grant, P.V. Eer, and R.A. Lester. Pandemic (H1N1) 2009 influenza community transmission was established in one Australian state when the virus was first identified in North America. PLoS ONE, 5(6), 2010.

[31] C. Fraser, C.A. Donnelly, S. Cauchemez, W.P. Hanage, M.D. Van Kerkhove, T.D. Hollingsworth, J. Grin, R.F. Baggaley, H.E. Jenkins, E.J. Lyons, T. Jombart, W.R. Hinsley, N.C.a Grassly, F. Balloux, A.C. Ghani, N.M. Ferguson, A. Rambaut, O.G. Pybus, H. Lopez-Gatell, C.M. Alpuche-Aranda, I.B. Chapela, E.P. Zavala, D. M. Espejo Guevara, F.

Checchi, E. Garcia, S. Hugonnet, and C. Roth. Pandemic potential of a strain of influenza A (H1N1): Early findings. Science, 324(5934):1557-1561, 2009.

[32] V.M. Carn. The role of dipterous insects in the mechanical transmission of animal viruses. British Veterinary Journal, 152(4):377-393, 1996.

[33] Stewart M. Gray and Nanditta Banerjee. Mechanisms of arthropod transmission of plant and animal viruses. Microbiology and Molecular Biology Reviews, 63(1):128-148, 1999.

[34] Sander P. van Noort, Claudia T. Codeco, Carl E. Koppeschaar, Marc van Ranst, Daniela Paolotti, M. Gabriela M. Gomes. Ten-year performance of influenzanet: ILI time series, risks, vaccine effects and care-seeking behaviour, Epidemiology 13 (2015) 28-36.

[35] J. Troko, P. Myles, J. Gibson, A. Hashim, J. Enstone, S. Kingdon, C. Packham, S. Amin, A. Hayward, and J. Nguyen Van-Tam. Is public transport a risk factor for acute respiratory infection? BMC Infectious Diseases, 11(16):1471-2334, 2011.

[36] Jason R. Andrews, Carl Morrow, and Robin Wood. Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. American Journal of Epidemiology, 177(6):556-561, 2013.

[37] David J Earn, Pejman Rohani, Benjamin M Bolker, and Bryan T Grenfell. A simple model for complex dynamical transitions in epidemics. Science, 287:667-670, 2000.

[38] Chris T Bauch and David J Earn. Transients and attractors in epidemics. Proceedings of the Royal Society of London B, 270(1524):1573-1578, 2003.

[39] Pejman Rohani, David J Earn, and Bryan T Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. Science, 286:968-971, 1999.