

A relationship between acute respiratory illnesses and weather

A. COSTILLA-ESQUIVEL¹, F. CORONA-VILLAVICENCIO¹,
J. G. VELASCO-CASTAÑÓN², C. E. MEDINA-DE LA GARZA^{2,3},
R. T. MARTÍNEZ-VILLARREAL^{3,4}, D. E. CORTES-HERNÁNDEZ^{2,3},
L. E. RAMÍREZ-LÓPEZ⁵ AND G. GONZÁLEZ-FARIAS^{1*}

¹ *Centro de Investigación en Matemáticas (CIMAT), Unidad Monterrey*

² *Universidad Autónoma de Nuevo León (UANL), Centro de Investigación y Desarrollo en Ciencias de la Salud*

³ *Universidad Autónoma de Nuevo León, Facultad de Medicina*

⁴ *Universidad Autónoma de Nuevo León, Centro Universitario de Salud*

⁵ *Coordinadora de Epidemiología, Jurisdicción número 4, Secretaría de Salud del Estado de Nuevo León*

*Received 12 September 2012; Final revision 30 April 2013; Accepted 10 July 2013;
first published online 2 August 2013*

SUMMARY

Weekly data from 7 years (2004–2010) of primary-care counts of acute respiratory illnesses (ARIs) and local weather readings were used to adjust a multivariate time-series vector error correction model with covariates (VECMX). Weather variables were included through a partial least squares index that consisted of weekly minimum temperature (coefficient = -0.26), weekly median of relative humidity (coefficient = 0.22) and weekly accumulated rainfall (coefficient = 0.5). The VECMX long-term test reported significance for trend (0.01 , $P = 0.00$) and weather index (1.69 , $P = 0.00$). Short-term relationship was influenced by seasonality. The model accounted for 76% of the variability in the series (adj. $R^2 = 0.76$), and the co-integration diagnostics confirmed its appropriateness. The procedure is easily reproducible by researchers in all climates, can be used to identify relevant weather fluctuations affecting the incidence of ARIs, and could help clarify the influence of contact rates on the spread of these diseases.

Key words: Climate – impact of, epidemics, modelling, respiratory infections, statistics.

INTRODUCTION

Acute respiratory illnesses (ARIs) stand out as the most common health complaint worldwide [1]. They were estimated to cause 4.2 million deaths and 97 million disability adjusted life years (DALYs) in 2004 [1, 2]. The usual syndrome consists of malaise, headache, myalgia, sore throat, coryza, fever, cough,

dyspnoea and additional signs of respiratory distress; clinical abnormalities generally arise in that order as severity of disease increases [3, 4].

Although bacteria may be present primarily or as a complication, it is viruses that cause ARIs most commonly (70–90% of cases). Over 200 different causal viruses have been identified, but most ARIs stem from rhinovirus, respiratory syncytial virus (RSV), coronavirus, adenovirus, metapneumovirus, and influenza or parainfluenza viruses [2, 5–8]. Public health impact of diseases from different viruses is distinct, and more attention has been devoted to those posing higher threats, such as influenza and

* Author for correspondence: Dr G. González-Farías, CIMAT, Centro de Investigación y Desarrollo en Ciencias de la Salud, Campus de la Salud, UANL, Ave Carlos Canseco s/n con Ave Gonzalitos, Col Mitras Centro Monterrey N.L. 64460, México.
(Email: farias@ciimat.mx)

RSV, through the former's potential for high-mortality pandemics, and the latter's more severe expression in children, often involving the lower airways.

In community settings most cases are mild, and the fraction of patients that seeks medical care usually receive symptomatic treatment without inquiry into the aetiological agent or agents, due to practical and financial constraints. Counts of ARI cases are kept and published in states that are members of the World Health Organization (WHO), and are used to report on global health and health-related economics [2]. These counts represent a proportion, likely to vary with age group, gender, and severity, of the real number of cases.

ARIs are related to weather; their name in many languages reflects this age-old observation (English: cold; Spanish: resfrío; German: Erkältung; French: froid). Indeed, counts of cases in temperate climates show peaks during their cold, dry, and low solar-radiation winters. On the other hand, tropical climates display an ARI incidence pattern with highs in the rainy season, during their warm summer [9–12]. It is of note that for some pathogens capable of causing ARIs, a biennial rhythmicity has been described (RSV, metapneumovirus, parainfluenza) in temperate regions; still, the seasonal variation of ARIs in general is annual [9, 13].

There have been efforts to quantify the relationship between ARIs and weather. Laboratory and observational studies suggest the existence of effects of low and high meteorological parameters for a variety of viruses on (1) the amount and viability of viruses shed by diseased individuals, (2) the probability of successful transmission by different mechanisms, and (3) host states that determine an individual's immunity and the group's vulnerability to an epidemic. Doyle & Cohen [3] observed that for viruses such as RSV or influenza, temperature, relative humidity, barometric pressure and ultraviolet radiation can explain 40% of the variance in the risk of ARI, and interpret this literature finding as related to the physical conditions favouring virus survival. Indeed, technical reports such as the one by Roberts & Kiang [12] on seasonal influenza transmission classify the meteorological factors recorded in the literature as those affecting virus survivorship (higher temperature, humidity, vapour pressure, and solar irradiance all decrease virus survival), those affecting transmission efficiency (decreased by higher temperature and vapour pressure, increased by higher rainfall, by ENSO, more air travel and

holidays, and with varying reports for humidity), and those affecting host susceptibility (decreased by higher sunlight exposure and with varying effects of nutrition). It is likely that the weather's influence on ARIs is mediated at least partially by effects on contact rates, and such effects may vary by type of climate and cultural group, as suggested by empirical observations [2, 6, 8, 14].

Our study used data from a primary-care population in the state of Nuevo León, in northeastern Mexico. Nearly four million inhabitants (about 80% of the state population in 2010) live in the metropolitan area of Monterrey, 537 m above sea level, and ~270 km north of the Tropic of Cancer. The climate is semi-arid (Köppen classification BSh), with a long hot season (>32 °C monthly average of daily maximum, individual readings as high as 45 °C, 2004–2010), correspondingly short cool and cold seasons, scarce precipitation (between 0.94 mm and 228.6 mm per day, mostly during the summer), and mild winters (between 4.5 °C and 7.4 °C annual average of daily minimum). During winter, the region experiences sudden changes in temperature, which may fall by 20 °C in a few hours, often accompanied by high humidity and generalized low-intensity rain.

This paper describes a vector error correction model with covariates (VECMX), that relates weekly counts of ARIs in a primary-care setting to weather variables in a specific climate, that of the northeast of Mexico, using data from 2004 to 2010. Models such as the one reported here may provide quantitative support for the role of contact rates in epidemic spread in naturalistic settings, by comparing the weather most likely to make people stay indoors in different cultures. If similar degrees of explanation are attained through the use of different weather variables in contrasting climates and peoples, human behaviour would be the parsimonious explanation. If, on the other hand, the same weather variables are related to increases in counts of cases worldwide, biophysical effects on viruses, hosts and groups should be attested to as more efficiently promoting the rapid dissemination of disease.

METHODS

Weekly ARI counts

The Universidad Autónoma de Nuevo León's (UANL) 11 primary-care clinics in the municipalities of Apodaca and Guadalupe, within the Monterrey

Table 1. ICD-10 codes and clinical diagnosis scored as acute respiratory illnesses

Code	Diagnosis
J00	Acute nasopharyngitis, common cold
J01.9	Acute sinusitis, unspecified
J02.9	Acute pharyngitis, unspecified
J04.0	Acute laryngitis
J04.1	Acute tracheitis
J04.2	Acute laryngotracheitis
J05.1	Acute epiglottitis
J06.0	Acute laryngopharyngitis
J06.9	Acute upper respiratory infection, unspecified
J20	Acute bronchitis
J20.9	Acute bronchitis, unspecified
J21	Acute bronchiolitis
J21.9	Acute bronchiolitis, unspecified
J11	Influenza, virus not identified

metropolitan area, provide medical care in low-income neighbourhoods. The data included weekly counts of the ICD-10 diagnoses listed in Table 1 (chronic ailments were excluded), as well as the age in years and the gender of the patient. Table 2 summarizes the demographical data of the sample.

Weather data

The study used information from the 'Monterrey' weather station (100° 16' 18" longitude, 25° 40' 57" latitude) for the same years, accessed through the National Water Commission (CONAGUA) website [15]. The readings consisted of mean temperature, maximum temperature, minimum temperature, barometric pressure, relative humidity, wind velocity, wind direction, and rainfall. After exploratory analysis, only the series of observations constituted by three parameters, i.e. minimum temperature during a given week, median of the observations of relative humidity during a given week, and rainfall accumulated during a given week, were retained in the model.

Statistical methods

Database conditioning

(a) *Case count database.* In order to correct for the bias introduced by non-biological variations in case counts, probably due to changes in the administrative organization of healthcare institutions affecting the number of persons served, such as the introduction of a new federal programme in 2004 known as Seguro Popular (Popular Insurance), the

Table 2. Gender and age group of patients in the acute respiratory illness counts (2004–2010)

Variable/value	<i>n</i>	Proportion	95% CI
Gender			
Male	13 717	0.43	0.42–0.44
Female	18 163	0.57	0.56–0.58
Age group (years)			
0–5	16 974	0.53	0.53–0.54
>5–15	6 576	0.21	0.20–0.21
>15–50	6 460	0.20	0.20–0.21
>50	1 870	0.06	0.056–0.061
Total	31 880		

CI, Rounded confidence interval.

counts of cases for 2004–2006 were scaled using state-wide data published by the State Health Ministry (Secretaría de Salud de Nuevo León), preserving the overall dynamics of the series. In addition, these Poisson counts were transformed to stabilize their variance and to smooth them, by $Y=2 [y^{1/2}]$ [16]. In addition, the demographic covariables were standardized, to render them less prone to noise from variation in the number of persons seeking attention related to administrative or political factors.

(b) *Meteorological database.* A small amount of observations (52/1092, 4.5%) were missing or clear errors (e.g. temperatures incompatible with life, relative humidity of zero accompanied by non-zero rainfall), and were substituted by estimates from a smoothing technique as proposed by Cleveland *et al.* in 1990 [17].

Model building

Weather variables modelling

As stated earlier, exploratory analysis led to inclusion in the model of three weather parameters: the minimum temperature during a week, the median of all readings of relative humidity in a week, and the weekly accumulated rainfall. As expected, variables in the meteorological database displayed linear and nonlinear dependencies, a situation known as multicollinearity. Regression procedures attempted with such data yield unstable and hard to interpret coefficients for the independent variables.

Therefore, we created a weather index through partial least squares (PLS) [18]. PLS is a technique that exploits the dependency structure in a set of variables

to determine a group of coefficients that best relate them to a dependent variable. PLS is similar to principal components analysis (PCA) and factor analysis (FA) in that they are all methods to reduce the dimensionality of a model by projecting a set of variables into a hyperplane that retains most of the information, but under different criteria. FA uses a model determined by the researcher, whereas PCA chooses a number of orthogonal dimensions based on maximizing the variance of the independent variables in each dimension, rendering it a more unspecific technique than PLS. PLS is also different in the fact that both the dependent and the independent variables are projected into a different space, with the aim of maximizing the covariance in them, as opposed to the unique variance maximization that occurs in PCA through the projection of one set of variables. The solution in PLS is a nonlinear function that is best suited for forecasting purposes. In our model, we were only interested in the coefficients of the solution produced, as a single index that captures the information in the three weather variables, and could be used in further statistical procedures [19, 20].

Vector error correction modelling

Multivariate time-series models undertake either the explanation or the prediction of the observations in one time-series by the dynamics of one or more other series that may summarize several variables. VECMXs are a subtype that seeks to do this by estimating: (1) long-term effects, (2) short-term or shock effects, and (3) an error term, which in a correctly specified model should be random noise. VECMXs allow the variables to ‘speak by themselves’, producing estimates of the long-term relationship between the time-series, decomposed in parameters of trend (how the series influence each other in the long term), seasonality (the time-frame in which oscillations around the trend are best understood), and autocorrelation [the extent of influence of the previous observation on subsequent one(s)] in the long-term relationship, and of the impact of sudden changes in the independent variables and covariates on the short-term relationship, including the speed with which the dependent series returns to the long-term relationship equilibrium [21]. In a problem such as the one we consider here, VECMX-relevant contribution is to determine if the dynamics present in the weather variables index are capable of reproducing the dynamics present in the ARI counts time-series.

Two covariates were included in the model. Age group and gender data were available for each case diagnosed, and were included since variations over time were observed on exploratory analyses. Once again, a PLS index for the proportions of each category allowed the addition of an age group covariate to the model. The gender covariate was added as the proportion of males. Since these primary-care clinics remained closed for 2 weeks for each Christmas and Easter holiday seasons, a covariable to adjust for these weeks, when zero cases meant that no patient requested attention, was also included. All these covariables were added because of particularities in the data, influenced by non-biological factors in this relatively small healthcare system, and thus will not be interpreted in the discussion of the model in epidemiological or biological terms.

We tested the fit of models with all combinations of three different parameters: trend (none, constant, positive), seasonality (3–52 periods), and lag (1–18 periods). It is of note that a VECMX model is able to provide spurious significance (regression artifacts), therefore the modelling procedure must include a diagnostic phase. In our model’s development, residual hypothesis-testing including autocorrelation tests supported the selection of a group of models. They were later tested for co-integration, believed to occur when there is a unique relationship in the series that does not depend on time, through the maximum eigenvalue long-term test [22–25]. From the models that remained after performing these diagnostics, we chose the one that maximized fit and plausibility.

Finally, we used an outlier Z test ($\alpha=0.10$) to detect atypical counts of cases that could not be explained by the model.

All the analyses were performed using the free open source statistical software R, specifically the packages ‘pls’ and ‘vars’ [26–28].

RESULTS

Weather index

The PLS solution to the problem of the combination of weather parameters that maximizes the covariance with the case counts, produced the coefficients described in Table 3 and equation (1).

$$\text{PLS weather index} = -0.26 \text{ minimum temperature} \\ \text{in the week} + 0.22 \text{ weekly median of relative} \\ \text{humidity} + 0.5 \text{ weekly accumulated rainfall.} \quad (1)$$

Table 3. *Partial least squares (PLS) coefficients for weather index and age group index*

PLS index	Variable	Coefficient
Weather	Minimum temperature in the week	-0.26
	Weekly median of relative humidity	0.22
	Weekly accumulated rainfall	0.50
Age group	0-5 years	0.24
	>5-15 years	0.43
	>15-50 years	0.54
	>50 years	0.33

The index is higher when the minimum temperature in the week is lower, and both weekly median of relative humidity and weekly accumulated rainfall are higher; the impact of rainfall is stronger than that of the other two parameters.

Covariates

Age group index

The PLS coefficients are described in Table 3. The resulting index assigned a lower weight to the transformed number of diagnosed cases in children aged ≤ 5 years, and assigned increasingly higher weights to those aged >50 , $>5-15$, and $>15-50$ years, in that order. Gender was added as the proportion of males in the counts of a given week. It is important to remember that we introduced this type of covariable only as an adjustment device, due to the sample characteristics, and not for an epidemiological reason.

$$\begin{aligned} \text{PLS age group index} = & (0.24) 0-5 \text{ years} \\ & + (0.43) > 5-15 \text{ years} + (0.54) > 15-50 \text{ years} \\ & + (0.33) > 50 \text{ years.} \end{aligned} \quad (2)$$

VECMX

In the selected model, the maximum eigenvalue test rejected the first null hypothesis, namely, the lack of a time-independent relationship, and did not reject the second null hypothesis, that there could be only one such relationship. It showed a slight positive trend of 0.01; the long-term equation was also significant, with a coefficient of 1.69 for the weather index ($P \leq 0.00$). In the short-term equation, the week of the year (4.42, $P=0.01$), having more older children and adults (1.50, $P \leq 0.00$), and less males (-2.43,

$P=0.04$), all had a statistically significant effect. The vector of short-term error corrections so formed was also significant (-0.67 , $P \leq 0.00$), and the overall degree of explanation was 76% (adj. $R^2=0.76$; for the most important coefficients, see Table 4). As stated earlier, the relevant information in a VECMX model for a problem like the one studied here is to determine the ability of one series, the weather index series in this case, to reproduce the dynamics of the dependent series, the ARI counts in this paper; this is shown graphically in Figure 1. (The full VECMX equation and table of coefficients is available upon request.)

Finally, hypothesis testing characterized the residuals as random noise, and values in the uppermost 10% of the observations were defined as atypical ($n=14$ weeks in the whole period, see Fig. 2).

DISCUSSION

We report a VECMX, using commonly kept data and free open source software, to help explain the dynamics of acute respiratory infection counts in terms of the weather in an urban community.

As mentioned, PLS maximizes the covariance in the weather series of observations and the ARI counts. The best possible fit between the transformed counts and the new variable, the PLS weather index that combines the three weather series, is given by awarding a -0.26 weight to the minimum temperature during a given week, a 0.22 weight to the median of relative humidity during a given week, and a 0.5 weight, the heavier influence, to accumulated rainfall during a given week. The signs of the coefficients imply the direction of the relationship: negative for minimum temperature means that the lowest temperatures are responsible for the relationship, while the positive signs for the other two parameters indicate that the highest readings are the ones that bear relation to increases in counts. This is in keeping with the literature on the most common viral agents [7-9, 11, 13, 29-33].

Both the gender and age-group PLS covariables render the dynamics of the influence of the weather index on the transformed counts clearer. It is known that young children are more vulnerable to ARIs and their complications than adults; and gender has been repeatedly proven not to affect the probability of ARI transmission ([3, p. 162, and references therein]). These covariables merely control for shifts in the covariances of the standardized demographic observations and the transformed counts of ARIs

Table 4. *VECMX adjusted parameters*

Type of relationship	Variable	Coefficient	<i>P</i> value
Long term	PLS weather index	1·60	0·00
	Trend	0·01	0·00
Short term	Error correction coefficient	-0·67	0·00
	<i>D</i> (PLS weather)	-0·66	0·11
	<i>D</i> (cases)	-0·05	0·43
	Seasonality	4·42	0·01
	PLS age group	1·50	0·00
	Gender	-2·43	0·04
	Holiday	0·17	0·79
Adjusted explanation	Adj. <i>R</i> ²	0·76	n.a.

VECMX, Vector error correction model with covariates; PLS, partial least squares; n.a., not available.

Bold values indicate statistical significance ($P \leq 0.05$).

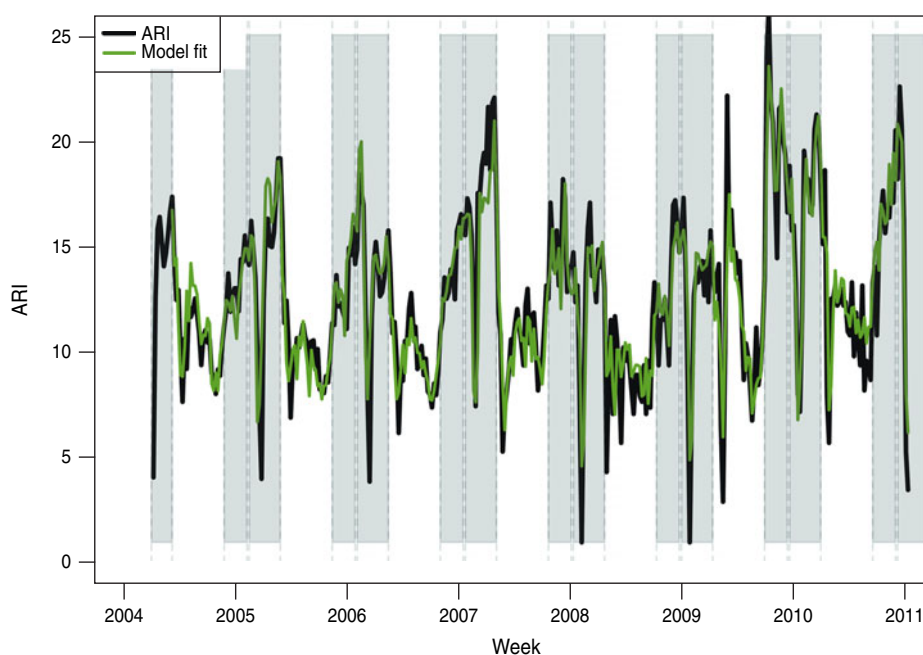


Fig. 1 [colour online]. Smoothed acute respiratory illness (ARI) series (black line) and model fit (grey line). Shaded areas in the background indicate the calendar autumn and winter seasons in each year.

along time-lines. The statistically significant trend in the model is quite small and thus lacks epidemiological importance; it may reflect demographical growth, somewhat blunted since the large non-biological fluctuations in the number of visits to the clinics were corrected for, as explained in the Methods section. The modelled relationship of weather to the ARI counts does have further impact on the short-term relationship, where only the last 3 weeks have significant coefficients, all with a negative sign. In other words, there were more cases in week T when

weeks T-1, T-2 and T-3 were characterized by high scores in the weather index, and when considering the dynamics of pairs of weeks, the previous one had a stronger influence ($T-3 > T-2 > T-1$), while this relationship diluted itself from T-4 onwards. However, the short-term relationship is characterized by a rather quick return to the long-term relationship, as can be inferred from the error correction coefficient of 0·67, meaning that 67% of the changes dissipate over a week. The proportion of the variability explained by the model is very satisfactory, 76%

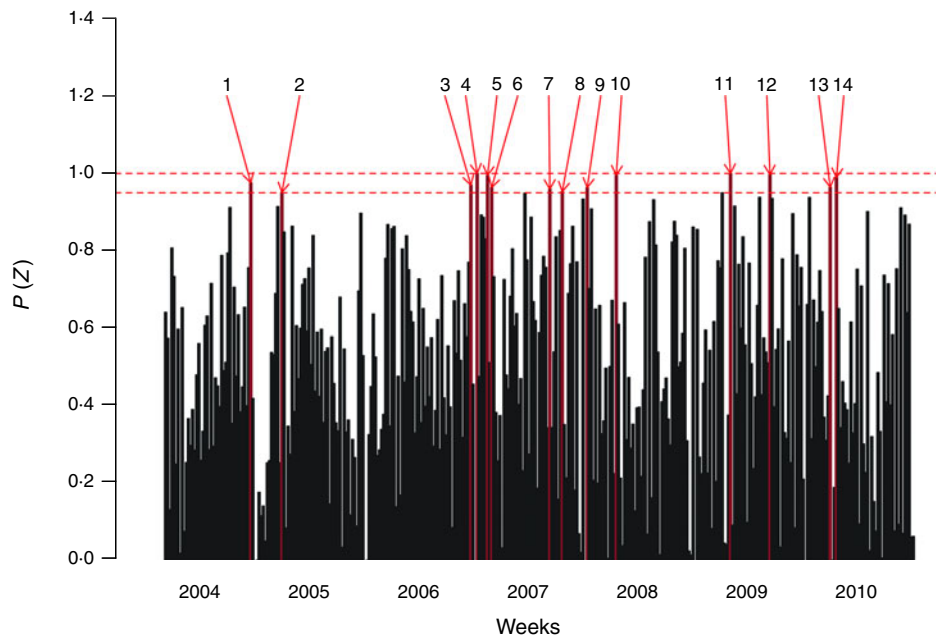


Fig. 2 [colour online]. Weeks with atypical counts (outbreaks). 1 = week 51, 2004; 2 = week 13, 2005; 3 = week 51, 2006; 4 = week 1, 2007; 5 = week 6, 2007; 6 = week 8, 2007; 7 = week 36, 2007; 8 = week 42, 2007; 9 = week 1, 2008; 10 = week 15, 2008; 11 = week 17, 2009; 12 = week 36, 2009; 13 = week 12, 2010; 14 = week 15, 2010.

(adj. $R^2=0.76$), the highest we were able to locate in the literature, perhaps owing to its multivariate (as opposed to multiple) nature.

The fitted model has two interesting advantages that, to our knowledge, previously reported models have not provided: (1) it lends itself to replication in different types of climate and different countries, due to its widely available types of data input, and (2) it has a low cost of implementation.

The first difference pertains to what at first may seem a lack of sophistication in the model, which does not consider the different pathogens involved. Indeed, successful efforts to relate weather parameters to the incidences of specific viral pathogens in ARIs have been previously reported [7–9, 11, 13, 29–33]. Laboratory and special population studies have shown effects of weather variables on the successful transmission by specific mechanisms for different viruses, affecting the length of time spent in the air for aerosol droplets, the survival of viruses on hands and objects, and the ability of the respiratory and immune systems to defend the individual from infection, among other parameters. Even though documented in laboratory conditions, the size of these effects in a naturalistic setting remains unclear.

It has been suggested that the weather's effect on contact rates is larger than the more biophysical virus-specific effects [2, 14]; further studies on the epi-

demiological dynamics of ARIs are needed to clarify the 'efficiency' of the effects to facilitate epidemics. In our opinion, models such as the one reported here may help to shed light on this issue. When adjusted for diverse climates and communities, they will determine if the kinds of weather that make a particular cultural group stay indoors, different for diverse latitudes and peoples, explain ARIs counts with a comparable degree of completeness. This possible result would weigh in favour of the increased contact rates explanation for the naturalistic spread of epidemics. On the other hand, the same modelling procedure may also be applied to carefully sampled counts of different aetiological agents, and thus inform specific prevention policy makers and their vaccination strategies [30].

The model we report does not imply causality. The reasons why cold rainy weeks are consistently associated with higher counts of cases 1–3 weeks later in a semi-arid climate in Mexico are not shown in these data. Yet, the described procedure presents an advantage to healthcare systems of any size, complexity, or global localization: an affordable way to compare the weather's influence on ARIs across the world using commonly kept data and software that is freely available through the internet.

Still another feature in the model, useful for decision-making authorities, is the detection of

atypically high numbers of cases. Epidemiological outbreak detection methods are in wide use, and the model provides a further check on this crucial public health function, which is so important because of the serious consequences of falsely reporting either an outbreak or its absence.

This set of procedures is easily reproducible for researchers throughout the world, and may add quantitative information about the means through which the weather influences the number of ARIs in different climates.

ACKNOWLEDGEMENTS

The authors thank Víctor Hugo Muñoz Sánchez, Ph.D. for his help in preparing the figures for this paper. This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT, grant no. 105657).

DECLARATION OF INTEREST

All of the authors hold academic or professional positions in public institutions and do not have any relationship with for-profit organizations, nor have any of them received an extraordinary stipend for the work hereby presented.

REFERENCES

1. **Monto AS.** Epidemiology of viral respiratory infections. *American Journal of Medicine* 2002; **112** (Suppl. 1): 4–9.
2. **Mathers C, Fat D, Boerma J.** *The Global Burden of Disease: 2004 Update*. Geneva: World Health Organization, 2008, pp. 54, 60.
3. **Doyle W, Cohen S.** Etiology of the common cold: modulating factors. In: Eccles R, Weber O, eds. *Common Cold*. Basel: Birkhäuser, 2009, pp. 149–186.
4. **Carrat F, et al.** Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American Journal of Epidemiology* 2008; **167**: 775–785.
5. **Nokso-Koivisto J, et al.** Viral etiology of frequently recurring respiratory tract infections in children. *Clinical Infectious Diseases* 2002; **35**: 540–546.
6. **Wong-Chew RM, et al.** Frequency of respiratory viruses and clinical characteristics in children attending a care centre in Mexico City [in Spanish]. *Salud Pública de México* 2010; **52**: 528–532.
7. **Tang JW, et al.** Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong. *Epidemiology and Infection* 2010; **138**: 226–235.
8. **Pica N, Bouvier NM.** Environmental factors affecting the transmission of respiratory viruses. *Current Opinion in Virology* 2012; **2**: 90–95.
9. **Tamerius J, et al.** Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental Health Perspectives* 2011; **119**: 439–445.
10. **du Prel JB, et al.** Are meteorological parameters associated with acute respiratory tract infections? *Clinical Infectious Diseases* 2009; **49**: 861–868.
11. **Soebiyanto RP, Adimi F, Kiang RK.** Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS ONE*. Published online: 1 March 2010. doi:10.1371/journal.pone.0009450.
12. **Roberts DO, Kiang RK.** Meteorological and environmental indicators in seasonal influenza transmission network. 2011 (http://gest.umbc.edu/student_opp/2010_sies_reports/DRoberts.pdf). Accessed 14 April 2013.
13. **Walton NA, et al.** Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *BMC Medical Informatics and Decision Making* 2010; **10**: 68.
14. **Brankston G, et al.** Transmission of influenza A in human beings. *Lancet Infectious Diseases* 2007; **7**: 257–265.
15. **Comisión Nacional del Agua – Estaciones Meteorológicas Automáticas.** (<http://smn.cna.gob.mx/emas/txt/>). Accessed 1 February 2012.
16. **Bickel P, Doksum K.** *Mathematical Statistics*. Oakland: Holden-Day, 1977, p. 32.
17. **Cleveland RB, et al.** STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* 1990; **6**: 3–73.
18. **Geladi P.** Herman Wold: the father of PLS. *Chemometrics and Intelligent Laboratory Systems* 1992; **15**: R7–R8.
19. **Garthwaite PH.** An interpretation of partial least squares. *Journal of American Statistical Association* 1994; **89**: 122–127.
20. **Janes KA, Yaffe MB.** Data-driven modelling of signal-transduction networks. *Nature Reviews Molecular Cell Biology* 2006; **7**: 820–828.
21. **Sims CA.** Macroeconomics and Reality. *Econometrica* 1980; **48**: 1–48.
22. **Johansen S.** Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 1991; **59**: 1551–1580.
23. **Johansen S.** Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 1988; **12**: 231–254.
24. **Johansen S, Juselius K.** Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 1990; **52**: 169–210.
25. **Johansen S.** Cointegration: overview and development. In: Andersen TG, Davis RA, Kreiss JP, Mikosch T, eds. *Handbook of Financial Time Series*. New York: Springer, 2009, pp. 671–693.
26. **The R Project for Statistical Computing** (<http://www.r-project.org/>). Accessed 11 April 2012.

27. **Pfaff B.** VAR, SVAR and SVEC models: implementation within R package vars. *Journal of Statistical Software* 2008; **27**: 1–32.
28. **Mevik BH, Wehrens R.** The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* 2007; **18**: 1–24.
29. **Chan PK, et al.** Seasonal influenza activity in Hong Kong and its association with meteorological variations. *Journal of Medical Virology* 2009; **81**:1797–1806.
30. **Urashima M, Shindo N, Okabe N.** A seasonal model to simulate influenza oscillation in Tokyo. *Japanese Journal of Infectious Diseases* 2003; **56**: 43–47.
31. **Chumkiew S, et al.** Climatic factors affecting on influenza cases in Nakhon Si Thammarat. *World Academy of Science, Engineering and Technology* 2007; **36**: 19–22.
32. **Viegas M, et al.** Respiratory viruses seasonality in children under five years of age in Buenos Aires, Argentina: a five-year analysis. *Journal of Infection* 2004; **49**: 222–228.
33. **Grassly NC, Fraser C.** Seasonal infectious disease epidemiology. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 2006; **273**: 2541–2550.