On the temporal association of outbreaks of Zika virus infection, newborn microcephaly, and Guillain-Barré syndrome

Leonelo E. Bautista

Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin at Madison, Madison, WI 53726-2397, USA

Correspondence to: Leonelo E. Bautista, MD, DrPH, Associate Professor. Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin at Madison, 610 Walnut Street, WARF 703, Madison, WI 53726-2397, USA. Email: lebautista@wisc.edu.

Comment on: Paploski IA, Prates AP, Cardoso CW, *et al*. Time Lags between Exanthematous Illness Attributed to Zika Virus, Guillain-Barré Syndrome, and Microcephaly, Salvador, Brazil. Emerg Infect Dis 2016;22:1438-44.

During the last couple of years Zika virus (ZIKV) infection has been considered an international public health emergency, due to possible but yet uncertain links with newborn microcephaly and Guillain-Barré syndrome (GBS). Concerns about these links originated from case reports and surveillance data suggesting an outbreak of ZIKV in Brazil was followed by outbreaks of GBS and microcephaly (1,2). A fundamental component of the public health response has been the implementation of a system for surveillance of ZIKV disease and its complications (3).

Seminal evidence of a temporal association between these outbreaks came from a study by investigators from the Centro de Pesquisas Gonçalo Moniz (CPGM), in Salvador, Bahia, the largest city in Northeastern Brazil (1). Though other studies also reported temporal associations between these outbreaks (4-6), only CPGM investigators have tested it and have made their data available to the public. They identified cases of acute exanthematous illness (AEI) through a sentinel surveillance system established by the Centers for Information and Epidemiologic Surveillance of Salvador (CIES). AEI surveillance started in April 2015, but medical records were reviewed to identify past cases up to February 15, 2015 (*Figure 1*). Cases were patients with rash, who did not meet diagnosis criteria for dengue, chikungunya, measles, or rubella, and were attributed to ZIKV because this was the main arbovirus circulating at that time (1,7).

In late May 2015 CIES started surveillance of patients hospitalized with "neurologic manifestations that might be linked to Zika", including GBS, with retrospective search of cases hospitalized during April-May. Authors provided no explicit reason why was GBS surveillance started, but this may have been triggered by an earlier unreplicated report of an increase in GBS cases during a Chikungunya outbreak in French Polynesia (8).

In October 2015 the CIES established a newborn microcephaly case-report system, following a report of an increase in cases in Pernambuco the previous month. Cases were defined as a head circumference <32 cm in full term and less than the third percentile of the Fenton curve in preterm newborns (1), and were searched retrospectively back to the start of the year. The first case was identified in mid-July, and no cases were identified before that time through the regular Brazilian Live Birth Information System (SINASC) (2). In November 19th 2015 the Brazilian Ministry of Health implemented an ad hoc public health surveillance system for microcephaly (2).

CPGM investigators used standard cross-correlations to identify the lag times (from 0 to 40 weeks) showing the highest correlations between weekly numbers of AEI cases and ensuing weekly numbers of cases of GBS and microcephaly. The strongest positive correlations occurred 5–9 weeks later for GBS and 30–33 weeks later for microcephaly. They argued these findings provided strong support for a positive association of AEI cases to ensuing cases of GBS and microcephaly.

CPGM investigators inferred an outbreak of microcephaly had occurred, based on the shape of the epidemic curve (1). Though the shape of the curve resembled that of a point outbreak, it could be also attributed to the intensity of
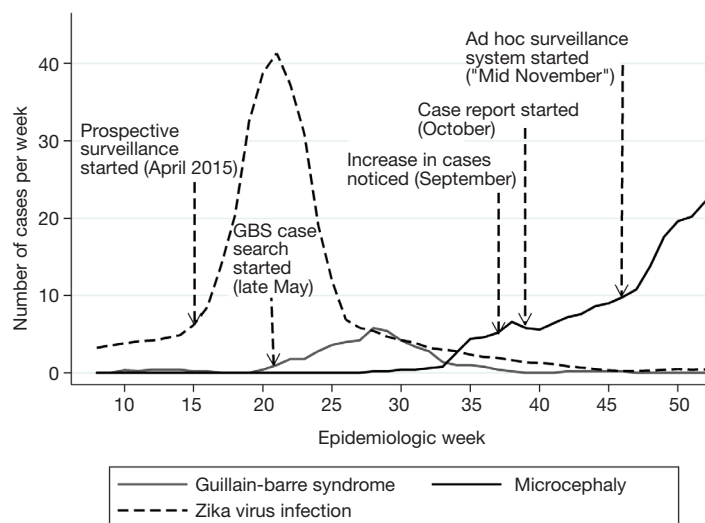
Figure 1 Epidemic curves of possible Zika virus infections (by 50), newborn microcephaly, and Guillain-Barre syndrome in Salvador, Brazil (2016).

surveillance efforts, and did not tell whether prevalence in 2015 was higher than expected. The prevalence of suspected microcephaly was 15.6 cases/1,000 newborns. Even in its peak (31.4 cases/1,000 in December 2015) the prevalence was below the expected 32.4 cases/1,000 newborns, inferred from the head size distribution in Brazilian newborns (normal, mean: 34.2 cm, standard deviation 1.2; see online Appendix for details on calculations and analyses) (9). Thus, CIES data provided no evidence of an outbreak of microcephaly in Salvador. An apparent outbreak may have resulted from comparing number of cases from the targeted-reactive CIES case-report system to those from the unspecific-regular SINASC in previous years. Indeed, SINASC traditionally had a low case yield (10), and detected no cases of microcephaly in Salvador during a period of 20 weeks in early 2015 (1), even though 462 cases were expected (see Appendix). Indeed, a low SINASC yield has been previously proposed by Brazilian experts on microcephaly as an explanation of the apparent outbreak (11).

The incidence of GBS in Salvador was 1.74/100,000 population (1). However, this rate pertained almost exclusively to the 33 weeks after the AEI peak on week 19, since only two out of 49 cases occurred in the 12 weeks period before that time (*Figure 1*). This was at variance with 57% of all AEI cases occurring before the AEI peak. If the rate observed after the AEI peak also applied to the whole year, the incidence would have been 2.74/100,000, well within the worldwide range of 0.6 to 4 per 100,000 (12,13). Thus, the apparent GBS outbreak in 2015 was likely due to

increased diagnosis, hospitalization, and case reporting after the start of surveillance of neurologic conditions.

CPGM investigators used time-lagged cross-correlations of moving averages to estimate induction times from ZIKV infection to GBS and microcephaly (1). Such an analysis is based on the assumption that causal links actually existed. Therefore, whatever the findings, they cannot be interpreted as evidence of assumed causal links, without incurring in the logical fallacy of circular reasoning.

It has been long established that cross-correlations of non-stationary time series are highly prone to bias (14,15). Outbreak are non-stationary series because their mean and variance change with time. For instance, the mean number of AEI cases increased with time until week 19 and decreased thereafter. Time series, like the outbreaks in question, may appear correlated, even if one does not causes the other, because they may share time-varying direct or mediated causes that confound their association (14,15). In other words, even if two outbreaks are cross-correlated, the distribution of one outbreak may not probabilistically depend on the distribution of the other, i.e., one outbreak may not independently predict the other outbreak. Also, the number of events in a time series are not independent. For instance, the number of cases of AEI in a given week would be closer to the number in the following week than to the number in more distant weeks. This autocorrelation is due to autocorrelation in the time-varying factors driving the series and must be accounted for during the analysis to get correct estimates of variability (16). Moreover, a time
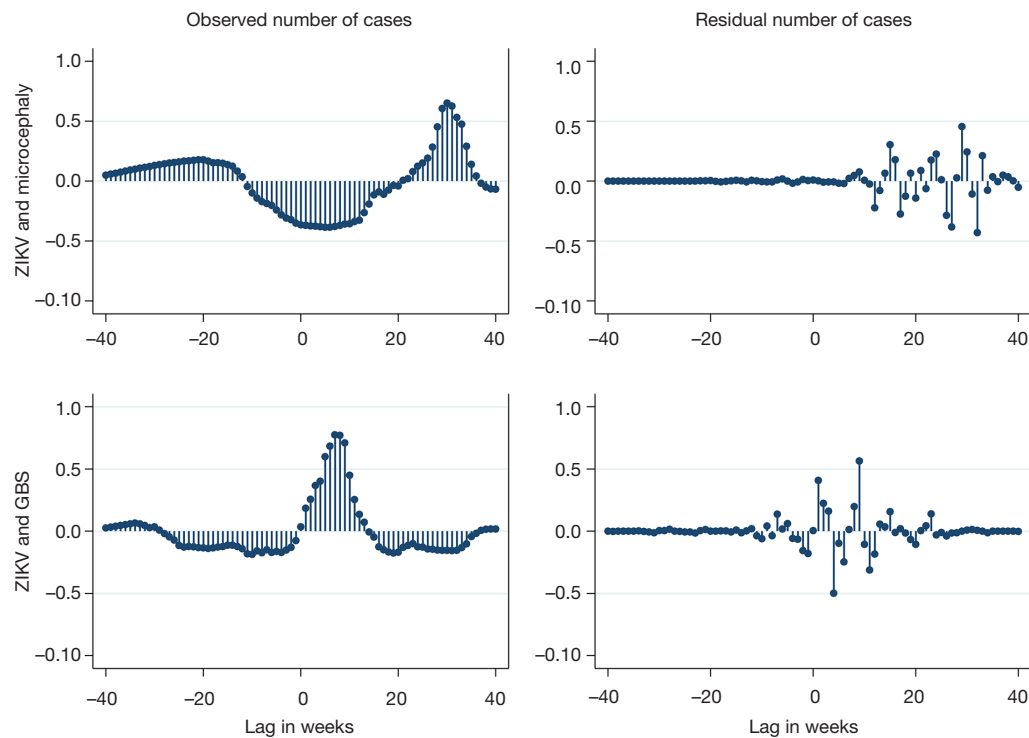
Figure 2 Cross-correlograms of weekly observed and residual number of cases of possible Zika virus infections (ZIKV), microcephaly, and Guillain-Barré syndrome (GBS), in Salvador, Brazil (2016).

series X could be deemed a cause of another time series Y if and only if predictions for Y differ when the history of X is taken into account and when it is not (17).

Induction times estimated by CPGM investigators were likely biased because they were based on statistical methods that assume constant mean and variance and no autocorrelation, the times corresponding to the highest correlations were automatically selected, and the correlations of moving-averages are more extreme than those of non-smoothed data.

I used a Poisson autoregressive model of first order to account for time "effects" and autocorrelation, and assessed whether the number of AEI cases independently predicted future changes in the number of cases of GBS and microcephaly (see Appendix) (18). I modeled the number of cases of each disease as a function of cubic polynomials splines of time (16,19). Then, I drew cross-correlograms of the observed number of cases, similar to those in the original study (1), and of the residual number of cases from the Poisson models.

Findings from the cross-correlations of the original data were consistent with those in the original study (*Figure 2*). However, after accounting for temporal fluctuations and autocorrelation there was no discernible pattern of correlation between the number of cases of AEI per week and the number of cases of GBS or microcephaly. Accordingly, the prevalence ratio (PR) for a change of 50 cases of AEI per week, for lag times (i.e., induction times) with the highest cross-correlations, were 1.01 [95% confidence interval (CI): 0.99, 1.03] for microcephaly (at 30 weeks), and 1.03 (95% CI: 0.98, 1.08) for GBS (at 7 weeks).

These findings suggest the apparent temporal correlations between the AEI outbreak and putative outbreaks of microcephaly and GBS (1), were likely a consequence of known biases in cross-correlation functions that do not account for time patterns and autocorrelation (14,15). Assigning a causal role to ZIKV infections cannot be based only on the observation of microcephaly and GBS outbreaks occurring after an AEI outbreak. Indeed, outbreaks could concur by chance or may share time-varying causes (i.e., confounders), such as time-related enhancement of surveillance efforts. They may also appear correlated due systematic errors in the data, such as non-independent errors in case ascertainment, or to the autocorrelated nature of the series. This type of causal inference based solely on temporal coincidence constitutes an example of the logical

fallacy *post hoc ergo propter* (i.e., after this, therefore because of it).

Another important report of a temporal association between ZIKV infection and microcephaly came from investigators from the Colombian National Institute of Health (INS), the US Centers for Diseases Control (CDC), and the Colombian Ministry of Health -MINSALUD (identified collectively as ICM) (6). The report was based on Colombian national surveillance data from August 9th 2015 to November 12th 2016, and included only suspected cases of ZIKV. The latter were defined as patients with fever and rash plus at least one of conjunctivitis, eye redness, itching, arthralgia, or malaise not explained by other diseases, who had been in a place below 2,200 m above sea level (20). Microcephaly was defined as head circumference below the third percentile for gestational age and sex (6).

ICM investigators reported a 4.5-fold increase in the prevalence of microcephaly in 2016 as compared to 2015 (from 2.1 to 9.6/10,000 births), as well as a temporal association between ZIKV infection and microcephaly, with a peak in the latter occurring approximately 24 weeks after the peak in ZIKV infections.

ICM's findings should be cautiously interpreted for various reasons. According to the Colombian INS protocol for surveillance of microcephaly, there were in average 140 cases of microcephaly per year during 2010–2015 (21). This estimate came from a mandatory registry of health care episodes, managed by MINSALUD. The registry gathers data on 80% of the Colombian population with health insurance, but it is known to capture <55% of all the episodes (22,23). Accounting for coverage, under-reporting, and duration or surveillance, 245 cases should have been expected in 2016. Thus, the prevalence in 2016 would have been 2.01 (95% CI: 1.72, 2.35) times higher than that in previous years. This increase that could be easily explained by intensified surveillance of microcephaly in 2016. Moreover, defining microcephaly as head circumference below the third percentile entailed an expected prevalence of 3% (6). Thus, the observed prevalence in 2016 (9.6/10,000) was 31 times lower than expected. This suggests that, in spite of the new guidelines (21), health care personnel continued reporting mostly cases of extreme microcephaly (head circumference ≤3 standard deviations from the mean) (24), and explains why the prevalence in Colombia was close to that estimated by passive surveillance in the United States (25). In summary, the available data gives no substantial support for the occurrence of an outbreak of microcephaly in Colombia in 2016.

Regarding the temporal association, a time series analysis following the approach described above, showed that the weekly number of ZIKV infections in Colombian pregnant women did not predict the prevalence of microcephaly weeks later (see Appendix). After accounting for autocorrelation and temporal fluctuations, there was no association between the weekly number of cases of ZIKV infections and the number of cases of microcephaly 19 weeks later, the time corresponding to the strongest correlation between these conditions (PR: 0.96; 95% CI: 0.92, 1.00).

Moreover, it is surprising that ICM investigators made no attempt to compare the prevalence of microcephaly in populations living above and below 2,200 m above sea level. Indeed, INS proposed residence below 2,000 m as a surveillance criterion to define ZIKV infections (21), and recognized that ZIKV transmission was unlikely and that over half the Colombian population lives above this altitude. Such a natural experiment should be free of biases attributable to time-varying factors, such as changes in the intensity of surveillance efforts, and could have shed light on the proposed ZIKV-microcephaly link.

Fortunately, the data provided in their supplemental table (6) made possible comparing the prevalence of microcephaly in Departments with some (374,488 births and 390 cases) and those with no population (123,514 births and 86 cases) living at or above 2,000 m (see Appendix). After adjusting for infant mortality and proportion of the population with health insurance in the Department, two likely surrogates of the strength of the surveillance system, the prevalence was only 1.62 (95% CI: 1.25, 2.10) times higher in populations living below than in those living above 2,000 m.

The above estimate may be biased towards the null due to non-differential misclassification resulting from the coarse grouping of Departments. However, a finer three-level stratification yielded similar findings (see Appendix). More important, a bias analysis assuming that sensitivity and specificity of the Department classification was non-differential regarding microcephaly status and varied from 60% to 90%, showed a highest PR of 2.84 (95% CI: 1.98, 4.10) for sensitivity of 80% and specificity of 60%. These findings could be reasonably attributed to a tradition of greater concern and better reporting of arbovirus diseases in low altitude areas, a hypothesis that could have been tested with available long term INS surveillance data on those diseases. If that hypothesis were discarded, one may be compelled to believe maternal ZIKV infection is unlikely to increase in the risk of microcephaly by more than 4 times.

The reports discussed in this commentary (1,6) exemplify the use of surveillance data to formulate causal link hypotheses and to guide health policies in a context of uncertainty. However, they also call attention to the need for caution in both instances. Surveillance data are usually not collected for the purpose of identifying causal links. In consequence, the potential for information, selection, confounding, and confirmation biases is exacerbated when surveillance data is used for this purpose. Also, political pressure to respond to ongoing outbreaks may make the need for public health action so strong that analyses of surveillance data and interpretation of findings could become mere exercises to confirm foregone conclusions. Some of these limitation may be mitigated by careful analyses and interpretation of surveillance data. This would be more feasible if institutions in charge of surveillance actively supported efforts to make data readily available in formats and by means that facilitate independent analyses by all stakeholders.

In summary, CPGM and ICM reports do not support the occurrence of microcephaly and GBS outbreaks in these populations or that the apparent outbreaks were linked to ZIKV infection outbreaks. Nevertheless, CPGM investigators should be highly commended for collecting unique and valuable data, for their transparency in describing how the data were collected, how cases were defined and identified, and, more important, for recognizing that surveillance data that informs public health and patient care decisions affecting millions of individuals should be made publicly available, so that all stakeholders have the opportunity and take responsibility for conducting their own assessments. ICM investigators should be also commended for collecting, processing and publishing some of their data, though more detailed data, and more information regarding data sources, and analytical decisions would have strengthened their report. Unfortunately, similar reports based on surveillance data have been notoriously lacking on these regards (2,4).

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Public Health and Emergency*. The article did not undergo external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jphe.2017.06.05). The author has no conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Paploski IA, Prates AP, Cardoso CW, et al. Time Lags between Exanthematous Illness Attributed to Zika Virus, Guillain-Barré Syndrome, and Microcephaly, Salvador, Brazil. Emerg Infect Dis 2016;22:1438-44.
2. Kleber de Oliveira W, Cortez-Escalante J, De Oliveira WT, et al. Increase in reported prevalence of microcephaly in infants born to women living in areas with confirmed Zika virus transmission during the first trimester of pregnancy - Brazil, 2015. MMWR Morb Mortal Wkly Rep 2016;65:242-7.
3. Pan American Health Organization. Guidelines for Zika virus disease and complications surveillance. Washington, D.C.: PAHO, 2016.
4. Dos Santos T, Rodriguez A, Almiron M, et al. Zika virus and the Guillain-Barré syndrome - case series from seven countries. N Engl J Med 2016;375:1598-601.
5. Machado-Alba JE, Machado-Duque ME, Gaviria-Mendoza A, et al. Diagnosis of Neurological Disorders and Zika virus epidemic in Colombia 2014 -2016. Int J Infect Dis 2016;51:133-4.
6. Cuevas EL, Tong VT, Rozo N, et al. Preliminary Report of Microcephaly Potentially Associated with Zika Virus Infection During Pregnancy-Colombia, January-November 2016. MMWR Morb Mortal Wkly Rep 2016;65:1409-13.
7. Cardoso CW, Paploski IA, Kikuti M, et al. Outbreak of

Exanthematous Illness Associated with Zika, Chikungunya, and Dengue Viruses, Salvador, Brazil. Emerg Infect Dis 2015;21:2274-6.

8. Oehler E, Fournier E, Leparc-Goffart I, et al. Increase in cases of Guillain-Barré syndrome during a Chikungunya outbreak, French Polynesia, 2014 to 2015. Euro Surveill 2015;20:30079.

9. Villar J, Cheikh Ismail L, Victora CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. Lancet 2014;384:857-68.

10. Pinto Júnior VC, Branco KM, Cavalcante RC, et al. Epidemiology of congenital heart disease in Brazil. Rev Bras Cir Cardiovasc 2015;30:219-24.

11. Lopez-Camelo JS, Orioli IM. ECLAMC Final Document. Summary and conclusions of documents 1-5. http://www nature com/news/zika-virus-brazil-s-surge-in-small-headed-babies-questioned-by-report-1 19259 [serial online] 2015. Available online: http://www.nature.com/polopoly_fs/7.33594!/file/NS-724-2015_ECLAMC-ZIKA%20VIRUS_V-FINAL_012516.pdf

12. Hughes RA, Rees JH. Clinical and epidemiologic features of Guillain-Barré syndrome. J Infect Dis 1997;176 Suppl 2:S92-8.

13. Hughes RA, Cornblath DR. Guillain-Barré syndrome. Lancet 2005;366:1653-66.

14. Yule GU. Why do we sometimes get nonsense-correlations between Time-Series? --a study in sampling and the nature of time-series. J R Stat Soc 1926;89:1-63.

15. Hoover KD. Nonstationary time series, cointegration, and the principle of the common cause. Br J Philos Sci 2003;54:527-51.

16. Bhaskaran K, Gasparrini A, Hajat S, et al. Time series regression studies in environmental epidemiology. Int J Epidemiol 2013;42:1187-95.

17. Granger CW. Some recent development in a concept of causality. J Econom. 1988;39:199-211.

18. Schwartz J, Spix C, Touloumi G, et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health 1996;50:S3-11.

19. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med 2000;19:1831-47.

20. Tolosa N. Protocolo de Vigilancia en Salud Pública. Enfermedad por virus Zika, 2016 [citado 21 Feb 2016]. Available online: https://www.google.com/ur

l?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjL4eHOkPrTAhVoyl QKHfWeDagQFggmMAA&url=http%3A%2F%2Fbvs.minsa.gob.pe%2Flocal%2FMINSA%2F3449.pdf&usg=AFQjCNFSvcopimfkut69OBi-xP47OA914g&sig2=BEf5WWUISMmIXc4DJY38gQ

21. Martínez MLO, Duran MEM, García OEP, et al. Protocolo de Vigilancia en Salud Pública. Vigilancia intensificada en salud pública de la microcefalia y otros defectos congénitos del Sistema Nervioso Central por virus Zika. 2016. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwihh66r_ovUAhUL wYMKHcFsBKYQFggmMAA&url=http%3A%2F%2Fbvs.minsa.gob.pe%2Flocal%2FMINSA%2F3449.pdf&usg=AFQjCNFSvcopimfkut69OBi-xP47OA914g

22. Bernal O, Forero JC, Villamil MadP, et al. Disponibilidad de datos y perfil de morbilidad en Colombia. Rev Panam Salud Pública 2012;31:181-7.

23. Ministerio de la Protección Social. República de Colombia. Demanda atendida en la población afiliada a los regímenes contributivo y subsidiado en el año 2009. Bogotá DC: Ministerio de la Protección Social. 2011. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiT zM_T_ovUAhUD_4MKHd1fApMQFggmMAA&url=h ttps%3A%2F%2Fwww.minsalud.gov.co%2FDocument os%2520y%2520Publicaciones%2FINFORMACI%25 C3%2593N%2520RIPS%25202009%2520-08202011. pdf&usg=AFQjCNFI5s2n3N3NGKE-jYG202tFMzOSRA&cad=rja

24. Ashwal S, Michelson D, Plawner L, et al. Practice parameter: evaluation of the child with microcephaly (an evidence-based review) report of the quality standards subcommittee of the American academy of neurology and the practice committee of the child neurology society. Neurology 2009;73:887-97.

25. Cragan JD, Isenburg JL, Parker SE, et al. Population-based microcephaly surveillance in the United States, 2009 to 2013: An analysis of potential sources of variation. Birth Defects Res A Clin Mol Teratol 2016;106:972-82.

This appendix provides additional details on calculations and analyses reported in the commentary and Stata code to replicate all results. In addition, an example of a nonsensical correlation is presented, to illustrate that spurious cross-correlations could be observed in unrelated time series and that the interpretation of those correlations depends on what we know and what we assume about the mechanisms generating both series.

Calculating the expected prevalence of microcephaly in Salvador, Bahia

This calculation is based on the distribution of head size among Brazilian newborn. According to the InterGrowth-21st study (9) the mean head circumference in Brazilian children is 34.2 cm, with a standard deviation of 1.2. The probability of having a head circumference <32 cm can be calculated from this distribution using a z test (see code for this calculation in item 1.0 of the accompanying Stata do file). This results in an expected prevalence of 32.8/10,000. This estimate only applies to full term newborns. However, 11.7% of Brazilian newborns are preterm (26). In the latter group the microcephaly was defined as a head circumference <3rd percentile of the Fenton curves (1), which corresponds to a prevalence of 30.0 cases/1,000 preterm newborns. Therefore, the expected prevalence in all newborns is a weighted average of the prevalences in full term and preterm newborns: $[(1-0.117) \times 32.8] + (0.117 \times 30) = 32.4$ cases/1,000 newborns.

Estimating the number of births and cases of microcephaly in Salvador in early 2015

No cases of microcephaly were found in Salvador before July 11th, 2015 (epidemiological week 27) and case search reached back to epidemiological week 7 (see *Figure S1* in original article) (1). Thus, no cases were found in a period of 20 weeks early in 2015. A total of 367 cases of microcephaly were found in the period of 33 weeks from July 2015 to February 2016. The prevalence during this period was 15.6/1,000 newborns. Therefore, the number of births in the period was $367/(15.6/1,000) = 23,526$. For a period of 20 weeks the number of births would be $23,526 \times (20/33) = 14,258$. Assuming the prevalence of 32.4/1,000 estimated above applies to the whole year, the number of expected cases of microcephaly would be $14258 \times (32.4/1,000) = 462$.

Estimating the incidence of Guillain-Barré syndrome (GBS) in Salvador in 2015

Paploski *et al.* (1) reported an incidence of GBS of 1.74 cases/100,000. This figure corresponds mostly to a period of 33 weeks, since only 2 cases of GBS were detected in the 12 weeks before the peak in AEI cases. However, 57% of all cases of AEI occurred during the period when only 2 cases of occurred, and cases were detected during this period by retrospective review of medical records, instead of prospectively reported. If one assumes that the incidence during the 33 weeks after the AEI peak applied to the whole year, the expected incidence in 2015 would have been $1.74 \times (52/33) = 2.74/100,000$. This incidence corresponds to an extreme scenario where the outbreak of Zika infections lasts the whole year and ika virus infection is actually associated with incidence of GBS.

Evaluating the temporal association between outbreaks in Brazil

Code for this analysis is provide in steps 3-6 of the Stata do file. Step 3 involves the creation of a dataset including the original data from Salvador, Bahia (3), available at http://wwwnc.cdc.gov/eid/article/22/8/16-0496-t1. It also includes the weekly number of cases of Entero-hemorrhagic *Escherichia coli* infection in Japan in 2016. The latter data is used for an example of a nonsensical correlation between two time series and were obtained from the Japanese National Institute of Infectious Diseases available at http://www.niid.go.jp/niid/en/survaillance-data-table-english.html?start=14. Epidemiologic curves are drawn in step 4. Dates of key events were added to the graph in the editorial using the Stata graph editor. Cross-correlations for the original number of cases and for residual number of cases are estimated and plotted in step 5. Code for autoregressive Poisson models to estimate the effect of the lagged number of cases of AEI on microcephaly and GBS is provided in step 6.

Figure S1 illustrate a nonsensical correlation between the weekly number of cases of *E Coli* infections in Japan in 2017 and the number of cases of microcephaly and GBS in Salvador in 2016. Were one unaware that data came from different times and places, upon looking at the cross-correlogram of the observed number of cases one may mistakenly conclude that cases of microcephaly increased significantly after a lag of 15 weeks after an increase in the cases of *E coli* (upper left panel). However, an analysis
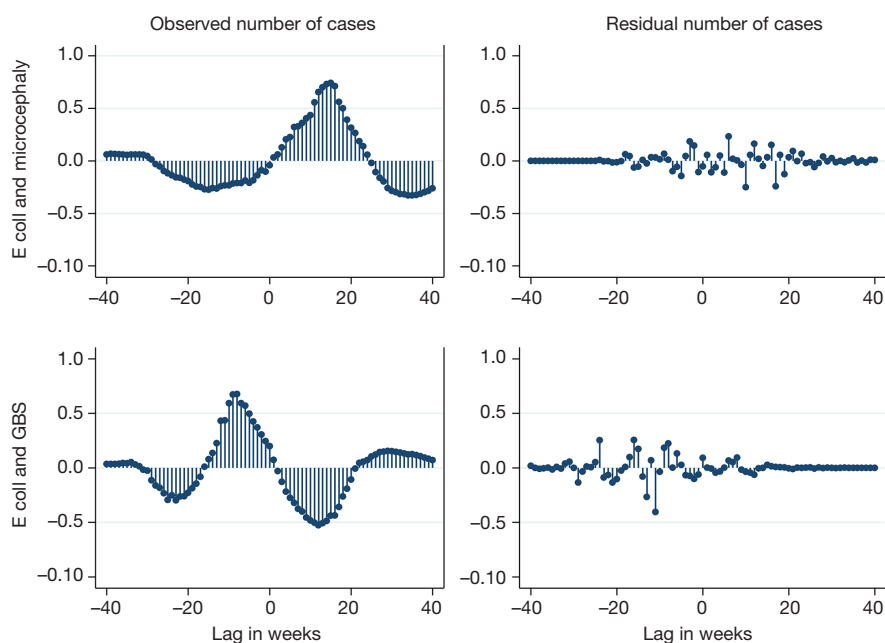
Figure S1 Cross-correlograms of weekly observed and residual number of cases of *E coli* infections in Japan (2017), and microcephaly and Guillain-Barré syndrome (GBS), in Salvador, Brazil (2016).

of the residual number of cases, after adjustment for time fluctuations and autocorrelation shows a different picture. Indeed, the cross-correlogram of the residual numbers of cases shows, as expected, that correlations between the two outbreaks (series) center around zero and that departures from a zero-correlation happens at random.

Comparing the observed and background cases of microcephaly in Colombia

Official Colombian data on the number of births during 2010–2015 and 2016 were obtained From: https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones/nacimientos. An official report from the Colombian INS indicates that the average number of cases of microcephaly in 2010–2015 was 140/year. After accounting for coverage, under-reporting, and duration or surveillance, the expected number of cases in 2016 was (140/0.55/0.80)×(40/52)=245 cases. The average number of births during the same period, in proportion to the number of weeks of surveillance in 2016 (40/52) was 510,926. The number of cases during the 40 weeks of surveillance in 2016 was 476 (6). The corresponding number of births was (641,493×40/52)=493,456. These figures were used to estimate the prevalence ratio. Instructions for this

calculation are provided in item 7 of the Stata code below.

Epidemiologic curves for maternal Zika virus infection and microcephaly in Colombia

Epidemiologic curves were drawn for a visual check of the accuracy of the process of recovering the number of cases from *Figure S1* of Cuevas *et al.* (6). Data were extracted from *Figure S1* using Engauge Digitizer 8.3 (27). The resulting Excel files were then converted to Stata data files for the purpose of the analysis. *Figure S2* shows that the time patterns in the number of cases were similar to those in *Figure S1* of Cuevas *et al.* (6). Departures from the original data were most likely random. See item 9 in the Stata do file. Excel data files are available from the author upon request.

Evaluating the temporal association between outbreaks in Colombia

Code for the cross-correlation and the autoregressive Poisson analyses are provided in item 10.1 and 10.2 of the Stata do file, respectively. *Figure S3* shows the findings from the cross-correlation analysis. In spite of the apparent high correlation of maternal Zika virus infections and
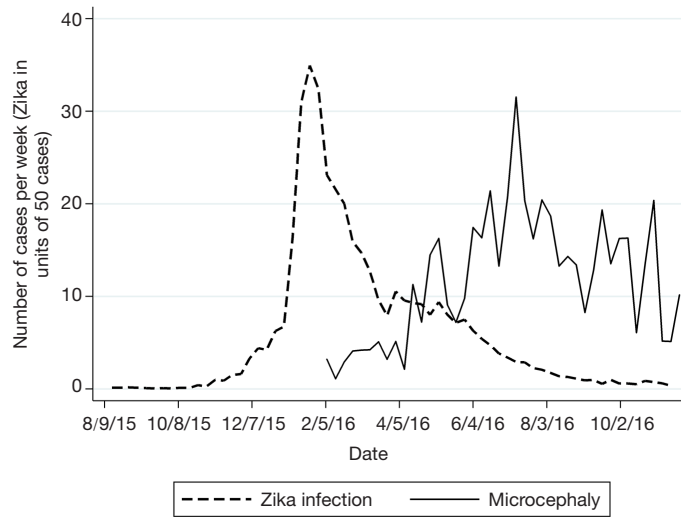
Figure S2 Epidemic curves of Zika virus infection in pregnant women and newborn microcephaly (based on captured data from original figure). Colombia, 2015–2016.
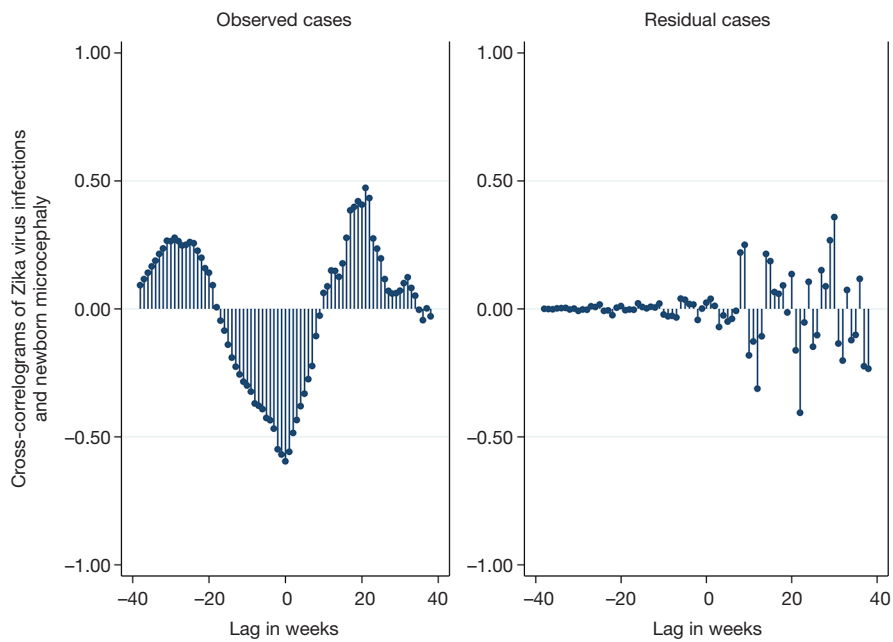
Figure S3 Cross-correlograms of weekly observed and residual number of cases of Zika virus infections and newborn microcephaly in Colombia (2015–2016).
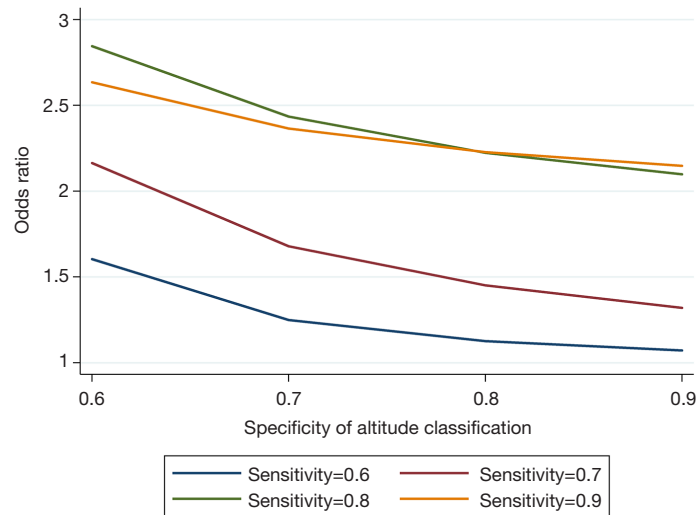
Figure S4 Odds ratios for altitude and microcephaly corrected for changing levels of misclassification of the exposure. Colombia (2015–2016).

microcephaly 19 weeks later (left panel), there was no discernible pattern in the correlation of the residual numbers of cases once time fluctuations and autocorrelation were accounted for (right panel). This indicates that maternal Zika virus infection did not probabilistically predict microcephaly.

Evaluating the impact of altitude on the prevalence of microcephaly in Colombia

See Stata code items 12.1 to 12.7. Excel data files available from author upon request. A Google maps key is required to extract the altitude of each Department. For details on how to get a key, go to: https://developers.google.com/maps/documentation/javascript/get-api-key. Altitude data are obtained through the Stata module "goelevation", which was kindly modified by its developer for the use of a Google maps key (28).

Findings from an analysis comparing Departments with all population leaving below and those with some population leaving above 2,000 m above sea level indicated a relative increase of 1.62 (95% CI: 1.25, 2.10) in the prevalence of microcephaly in the former group. Finding from an analysis with Departments grouped in those with 0%, >0 to <60%, and ≥60% of the population leaving above 2,000 m were consistent with those from the previous analysis. The prevalence was increased 1.68 (95% CI: 1.28, 2.20) times and

1.51 (1.12, 2.04) times in the population from Departments with >0 to <60% and ≥60% of the population living below 2,000 m, respectively. These groups were combined, since the prevalence ratios were very similar.

The classification of Departments by altitude in this analysis is based on the proportion of the population living above or below 2,000 m. Therefore, it is possible that some Departments were misclassified. However, the misclassification should be non-differential, because the proportion of the population living at high altitude is not a consequence of the prevalence of microcephaly in the Department. To account for misclassification of the exposure, I used maximum likelihood logistic regression with exposure predictive value weighting, as proposed by Lyles and Lin (29). I used values of sensitivity and specificity of the exposure classification ranging from 60% to 90%. Lower values (closer to 50%) would indicate a random assignment of exposure, while higher values (closer to 100%) would indicate a perfect classification. The last two scenarios are unlikely, because individuals move freely between low and high altitude areas, and there is evidence that transmission of ZIKV could occur in areas of high elevation, depending on latitude, climate, and connectivity of areas (30). *Figure S4* shows that correcting for non-differential misclassification of Department would increase the prevalence ratio, but in average the increase would be less than three-fold.

STATA CODE FOR THE ANALYSIS.

/************************DATA FROM SALVADOR, BAHIA, BRAZIL*****************************/
/**/

/*1.0 CALCULATING THE EXPECTED PREVALENCE OF MICROCEPHALY IN SALVADOR, BAHIA*/
/*Note: According to Villar J et al. Lancet 2014; 384: 857–68 (INTERGROWTH-21st) the mean head
circumference in Brazilian children is 34.2 cm, with a standard deviation of 1.2. The probability of having a
head circumference <32 cm can be calculated from this distribution using a z test.*/

```
scalar z=(31.99-34.2)/1.2
dis as text "Expected prevalence of HC<32 cm in full terms: " as res %4.2f (1-normal(abs(z)))*1000 " /1,000"
dis as text "Expected prevalence of microcephaly in all newborns: " as res %4.2f (((1-normal(abs(z)))*1000)*(1-
0.117)) + (30.0*0.117)
```

**/*2.0 CALCULATING THE INCIDENCE OF GBS, UNDER THE ASSUMPTION OF LOW DETECTION RATE IN EARLY
2015*/**
/*Note: Paploski et al. (Emerg Infect Dis. 2016;22(8)), reported an incidence of Guillain-Barré syndrome of
1.74 cases/100,000, but this figures corresponds mostly to a period of 33 weeks, since only 2 cases were
detected in the rest of the year. However, 57% of all cases of AEI ocurred during the period when only 2 cases
of occurred, a period when cases were detected through retrospective review of medical records, instead of
prospectively reported.*/

```
scalar prevgbs=(1.74*52/33)
dis as text "Incidence of GBS in 2015: " as res %4.2f prevgbs "/100,000"
```

/*3.0 CREATING THE DATASET*/
/*Note: In this step a Stata dataset is created. The dataset includes the data from the original article (Paploski
et al. Emerg Infect Dis. 2016;22(8)), which are available at http://wwwnc.cdc.gov/eid/article/22/8/16-0496-t1.
It also includes the weekly number of cases of Entero-hemorrhagic *Escherichia coli* infection in Japan in 2016.
The latter data were obtained from the Japanese National Institute of Infectious Diseases at
//http://www.niid.go.jp/niid/en/survaillance-data-table-english.html?start=14.*/

```
clear
input   week   year   aei    gbs    micro   ecoli
        7      2015   161    0      0       9
        8      2015   195    0      0       13
        9      2015   216    1      0       10
        10     2015   245    0      0       6
        11     2015   228    1      0       7
        12     2015   242    0      0       10
        13     2015   288    0      0       7
        14     2015   543    0      0       15
        15     2015   840    0      0       7
        16     2015   1585   0      0       18
        17     2015   1861   0      0       20
        18     2015   3301   0      0       20
        19     2015   2105   2      0       30
        20     2015   1486   3      0       23
```

21	2015	551	4	0	36
22	2015	279	0	0	35
23	2015	366	5	0	65
24	2015	294	6	0	87
25	2015	229	5	0	101
26	2015	289	5	0	118
27	2015	177	8	1	110
28	2015	179	3	0	119
29	2015	181	0	1	92
30	2015	150	1	0	123
31	2015	118	2	1	145
32	2015	127	1	2	142
33	2015	121	1	9	238
34	2015	72	0	10	151
35	2015	77	0	1	232
36	2015	86	0	4	191
37	2015	54	0	9	119
38	2015	50	0	5	77
39	2015	55	0	9	66
40	2015	43	0	5	65
41	2015	12	1	8	54
42	2015	10	0	11	79
43	2015	2	0	10	80
44	2015	14	0	11	69
45	2015	21	0	9	76
46	2015	16	0	13	44
47	2015	17	0	26	36
48	2015	31	0	29	27
49	2015	35	0	21	16
50	2015	4	0	12	22
51	2015	21	0	23	49
52	2015	9	0	14	18
1	2016	.	0	20	.
2	2016	.	0	21	.
3	2016	.	0	13	.
4	2016	.	0	17	.
5	2016	.	0	9	.
6	2016	.	0	12	.
7	2016	.	0	10	.
8	2016	.	0	2	.
9	2016	.	0	6	.
10	2016	.	0	3	.

end

/*3.1 MAKING CHANGES TO THE DATASET AND CREATING NEW VARIABLES*/
drop if year==2016 /*Dropped because there are no AEI observations in 2016*/
label var week "Epidemiologic week"
label var year "Year"
label var aei "Acute exhantematous illness"
label var gbs "Guillain-Barre syndrome"

```
label var micro "Microcephaly"
label var ecoli "E coli Japan 2017"
label var gbs "Guillain-Barre syndrome (GBS)"
qui gen aei50=aei/50          /*This is AEI number of cases in units of 50 cases*/
label var aei50 "AEI in units of 50 cases"
qui gen ecoli10=ecoli/10      /*This is E coli cases in units of 10 cases*/
label var ecoli10 "E coli in units of 10 cases"


/*3.2  SMOOTHING THE TIME SERIES USING MOVING AVERAGES*/
tsset week
tssmooth  ma aeima = aei50, window(5)
label var aeima "MA Zika infections"
tssmooth  ma gbsma = gbs, window(5)
label var gbsma "MA Guillain-Barre syndrome"
tssmooth  ma microma = micro, window(5)
label var microma "MA microcephaly"
tssmooth  ma ecolima = ecoli10, window(5)
label var ecolima "MA E coli"
capture log close
save jphe1.dta, replace


/*4.0 CREATING SMOTHED EPIDEMIOLOGIC CURVES, GRAPH 1; 5-WEEK SMOOTHING*/
/*Note: Dates of key events were added to the graph in the commentary using the Stata graph editor. Also, a
curve for E coli infections in Japan has been added to the curve generated by the code below. This curve is
presented as an example of a "nonsensical correlation". Cases of E coli in Japan in 2016 cannot influence the
incidence of microcephaly in Salvador in 2015. However, had one had no information on how the data were
generated, by looking at the graph one could have postulated that E coli caused microcephaly after a lag of 15
weeks, a hypothesis similar to that formulated for AEI, Guillain-Barré syndrome, and microcephaly.*/

capture program drop doepicurves
qui program define doepicurves
use jphe1.dta, clear
tsset week
label var gbsma "Guillain-Barre syndrome"
label var microma "Microcephaly"
label var ecolima "E coli infections"
label var aeima "Zika virus infection"
graph twoway  (line gbsma week, lp(solid)  color(black)  lwidth(medthick) msymbol(circle) msize(small))    ///
              (line microma week,  lp(solid) lwidth(medthick) color(black))    ///
              (line ecolima week,  lp(dash)  lwidth(medthick) color(black))     ///
              (line aeima week,   lp(dash)  lwidth(medthick) color(black)  sort),  ///
              yscale(range(0 45) axis(1)) ylabel(0 10 20 30 40, axis(1))   ///
              tline(15 21 37 39 46, lp(dash) lc(black))  ///
              ytitle("Number of cases", axis(1)) xlabel(10 15 20 25 30 35 40 45 50)     ///
              ti("Figure 1. Epidemic curves of possible Zika virus infections (by 50),"   ///
              "newborn microcephaly, and Guillain-Barre syndrome in Salvador,"     ///
              "Brazil (2016), and E. coli infections (by 10) in Japan (2017).",              ///
               justification(left) size(medium) color(black))
end
doepicurves
```

/*5.0 CALCULATING ANG PLOTTING CROSS-CORRELATIONS BEFORE AND AFTER ADJUSTING FOR TIME TRENDS AND AUTOCORRELATION.*/

/*Note: I generated different cubic polynomial splines for each of AEI, GBS, and microcephaly taking into account there were only 46 observations and the number of cases for GBS and microcephaly was small for some weeks*/

/*5.1 GRAPHICALLY VERIFYING THAT THE SMOOTHING OF CASES PER WEEK USING CUBIC POLYNOMIAL SPLINES OF WEEK FITTED THE OBSERVED DATA WELL*/

```
capture program drop splfit
qui program define splfit
use paploski5115_jphe.dta, clear
qui frencurv, xvar(week) gen(splaei)  power(3) refpts(5 10 15 20 25 30)  /*This corresponds to cubic equations
for refpts+2 curve segments with only 46 total observations*/
qui glm aei50 splaei2-splaei7, fam(poisson) scale(x2) nocons
qui predict aei50xb, xb              /*linear predictor of aei50*/
qui gen aei50xbnumber =exp(aei50xb)    /*predicted number of cases of AEI*/
label var aei50xbnumber "Predicted number of cases"
qui gen resaei=aei50-aei50xbnumber     /*residual number of AEI cases after accounting for time effects*/
graph twoway (scatter aei50 week) (line aei50xbnumber week), ti("Figure_aei. Epidemic curves of observed
and predicted cases of possible" "Zika virus infections (by 50), in Salvador, Brazil (2016).", justification(left)
size(medium) color(black))

qui frencurv, xvar(week) gen(splgbs)  power(3) refpts(5 10 15 20 25 30 40)
qui glm gbs splgbs*, fam(poisson) scale(x2) nocons
qui predict gbsxb, xb
qui gen gbsxbnumber =exp(gbsxb)
label var gbsxbnumber "Predicted number of cases"
qui gen resgbsi=gbs-gbsxbnumber
graph twoway (scatter gbs week) (line gbsxbnumber week), ti("Figure_gbs. Epidemic curves of observed and
predicted cases of" "Guillain-Barre syndrome, in Salvador, Brazil (2016).", justification(left) size(medium)
color(black))

qui frencurv, xvar(week) gen(splmicro)  power(3) refpts(20 30 35 40 45 50)
qui glm micro splmicro*, fam(poisson) scale(x2) nocons
qui predict microxb, xb
qui gen microxbnumber =exp(microxb)
label var microxbnumber "Predicted number of cases"
qui gen resmicroi=micro-microxbnumber
graph twoway (scatter micro week) (line microxbnumber week), ti("Figure_micro. Epidemic curves of observed
and predicted cases of" "newborn microcephaly, in Salvador, Brazil (2016).", justification(left) size(medium)
color(black))

qui frencurv, xvar(week) gen(splecoli)  power(3) refpts(5 10 15 20 25 30 40)
qui glm ecoli splecoli*, fam(poisson) scale(x2) nocons
qui predict ecolixb, xb
qui gen ecolixbnumber =exp(ecolixb)
label var ecolixbnumber "Predicted number of cases"
qui gen resecoli=ecoli-ecolixbnumber
graph twoway (scatter ecoli week) (line ecolixbnumber week), ti("Figure_ecoli. Epidemic curves of observed
and predicted cases of E coli" "infections, in Japan (2017)", justification(left) size(medium) color(black))
end
```

splfit

/*5.2 OBTAINING CROSS-CORRELATIONS BEFOR AND AFTER ACCOUNTING FOR THE EFFECT OF TIME.
Brazilian data*/
/*Note: Cross-correlograms for E coli infections are included in this part of the analysis. Looking at Figure 3 one can see that a series from a different place and a future time (E coli) results in cross-correlation patterns that are similar to those from a series (AEI) that happened before and at the same place as the GBS and the microcephaly series. However, when the "effect" of time is taken into account the nonsensical correlations with E coli are no longer present. Therefore, even if the postulated association between times series makes sense, without accounting for the non-stationary, autocorrelated nature of the series, it is impossible to infer that the series were actually associated.*/

```
capture program drop rescorr
qui program define rescorr
expand 2, gen(expanded) /*Expands the data set to get cross correlations up to 40 weeks. Dropping
expanded=1 returns the original data*/
qui xcorr aei micro, saving(aei_micro.gph, replace) lags(40) ytitle("ZIKV and microcephaly") ti("Observed
number of cases", justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("")
ylabel(,format(%4.1fc)) msize(small)
qui xcorr resaei resmicro, saving(resaei_resmicro.gph, replace)  lags(40) ytitle("Zika and microcephaly")
ti("Residual number of cases", justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("") ytitle("")
ylabel(,format(%4.1fc)) msize(small)
qui xcorr aei gbs, saving(aei_gbs.gph, replace)  lags(40) ytitle("ZIKV and GBS") ti("", justification(left)
size(medium) color(black)) ylabel("", axis(2)) xtitle("Lag in weeks") ylabel(,format(%4.1fc)) msize(small)
qui xcorr resaei resgbs, saving(resaei_resgbs.gph, replace)  lags(40) ytitle("ZIKV and GBS") ti("",
justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("Lag in weeks") ytitle("")
ylabel(,format(%4.1fc)) msize(small)
qui xcorr ecoli micro, saving(ecoli_micro.gph, replace)  lags(40) ytitle("E coli and microcephaly") ti("Observed
number of cases", justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("")
ylabel(,format(%4.1fc)) msize(small)
qui xcorr resecoli resmicro, saving(resecoli_resmicro.gph, replace)  lags(40) ytitle("E coli and microcephaly")
ti("Residual number of cases", justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("") ytitle("")
ylabel(,format(%4.1fc)) msize(small)
qui xcorr ecoli gbs, saving(ecoli_gbs.gph, replace)  lags(40) ytitle("E coli and GBS") ti("", justification(left)
size(medium) color(black)) ylabel("", axis(2)) xtitle("Lag in weeks") ylabel(,format(%4.1fc)) msize(small)
qui xcorr resecoli resgbs, saving(resecoli_resgbs.gph, replace)  lags(40) ytitle("E coli and GBS") ti("",
justification(left) size(medium) color(black)) ylabel("", axis(2)) xtitle("Lag in weeks") ytitle("")
ylabel(,format(%4.1fc)) msize(small)

gr combine aei_micro.gph resaei_resmicro.gph aei_gbs.gph resaei_resgbs.gph,  ///
saving(Figure2_JPHE.gph, replace)  ///
ti("Figure 2. Cross-correlograms of weekly observed and residual number of" ///
"cases of possible Zika virus infections (ZIKV), microcephaly, and Guillain-" ///
"Barre syndrome (GBS), in Salvador, Brazil (2016).", ///
justification(left) size(medsmall) color(black))

gr combine ecoli_micro.gph resecoli_resmicro.gph ecoli_gbs.gph resecoli_resgbs.gph, ///
saving(Figure3_JPHE.gph, replace) ///
ti("Appendix Figure 1. Cross-correlograms of weekly observed and residual" ///
"number of cases of E coli infections in Japan (2017), and microcephaly" ///
```

"and Guillain-Barre syndrome (GBS), in Salvador, Brazil (2016).", ///
justification(left) size(medsmall) color(black))

```
drop if expanded==1
drop expanded
end
rescorr
```

/*6.0 FITTING AUTOREGRESSIVE POISSON MODEL FOR THE LAG TIME WITH THE HIGHEST CROSS-CORRELATION*/

/*Note: The code below estimates the effect of AEI, in units of 50 cases, on the number of cases of microcephaly 30 weeks later, and on the number of cases of GBS 7 weeks later. This time points corresponds to the induction periods inferred from the maximum values of the cross-correlation functions of the original data (i.e. without accounting for time effects and autocorrelation).*/

```
capture program drop maxeffects
qui program define maxeffects
use paploski5115_jphe.dta, clear
frencurv, xvar(week) gen(spl)  power(3) refpts(25 35 45)
qui for num 0/35: qui gen aei50lagX = aei50[_n-X]
qui for num 0/35: qui gen Daei50lagX = D.aei50lagX
qui for num 0/35: qui gen ecolilagX = ecoli[_n-X]
qui for num 0/35: qui gen DecolilagX = D.ecolilagX

qui arpois micro Daei50lag30 spl2-spl5, prev delete ar(1)
qui lincom X1
scalar lag=30
scalar rr=exp(r(estimate))
scalar lbrr=exp(r(estimate)-1.96*r(se))
scalar ubrr=exp(r(estimate)+1.96*r(se))
dis as text _col(2) "Rate ratio of microcephaly per change of 50 cases in AEI 30 weeks before: "  as res _col(80)
`lag' _col(15) %6.4f rr " (" %6.4f lbrr ", " %6.4f ubrr ")"

qui arpois gbs Daei50lag7, prev delete ar(1)
qui lincom X1,
scalar lag=7
scalar rr=exp(r(estimate))
scalar lbrr=exp(r(estimate)-1.96*r(se))
scalar ubrr=exp(r(estimate)+1.96*r(se))
dis as text _col(2) "Rate ratio of GBS per change of 50 cases in AEI 7 weeks before: "   as res _col(10) `lag'
_col(80) %6.4f rr " (" %6.4f lbrr ", " %6.4f ubrr ")"
end
maxeffects
capture log close
```

/*****************************DATA FROM COLOMBIA*********************************/
/**/

/*7.0 ESTIMATING THE PREVALENCE RATIO OF MICROCEPHALY COMPARING 2016 TO 2010-2015*/
```
local caseback=round((140/0.55/0.80)*(40/52) ,1)
```

```stata
local birth10_15=round(((((654627+665499+676835+658835+669137+660287)/6)*40/52),1) /*From:
https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones/nacimientos*/
local case2016=476
local births2016=round((641493*40/52),1) /*Same source as above*/
local ncaseback=`birth10_15'-`caseback'
local ncase2016=`births2016'-`case2016'
csi `case2016' `caseback' `ncase2016' `ncaseback'
```

```stata
/*8.0 EVALUATION WHETHER THE OUTBREAKS IN COLOMBIA WERE TEMPORALY ASSOCIATED*/
/*8.1 Creating the data set for the number of cases per week*/
/*Note: Data for this evaluation were extacted from Figure 1 (Cuevas EL: MMWR /
        December 9, 2016 / Vol.65) using Engauge Digitizer 8.3. Excel files
        are available from the author upon request.*/
clear
import excel zikv1-Col.xlsx, first
replace week=round(week,1)
sum zika if week>=27 & week<=29 /*Point was not tagged in Engauge Digitizer*/
set obs 66
replace week = 28 in 66
replace zika = r(mean) in 66
sort week
tsset week
tsline zika
gen zika50=zika/50
save t1, replace

clear
import excel micro1-Col.xlsx, first
replace week=_n+26
tsset week
tsline micro
sort week
merge 1:1 week using t1
drop _merge
sort week
tsset week
label var micro "Microcephaly"
label var zika50 "Zika infections"
label var week "Epidemiological week"
gen week2=.
replace week2=week+30 if week<23
replace week2=week-22 if week>=23
label var week2 "Epidemiological week"
gen year=.
replace year=2015 if week<23
replace year=2016 if week>=23
label var year "Year"
egen wy = concat(year week2), p(w)
epiweek2 wy, s(from) e(to)
gen mid= (to- from)+from
```

```
format mid %tdnn/dd/YY
label var mid "Date"
save col_outbreak, replace
```

/*9.0 CHECKING THAT EPIDEMIOLOGIC CURVES FROM OBSERVED DATA (FIGURE 1, CUEVAS ET AL.) AND FROM DATA EXTRACTED USING ENGAUGE DIGITIZER WERE CONSISTENT*/

```
twoway line zika50 micro mid, ti("Appendix Figure 2. Epidemic curves of Zika virus infection in pregnant" ///
 "women and newborn microcephaly. Colombia, 2015-2016.", justification(left) size(medsmall) color(black)) ///
    tlabel(09aug2015 (60) 02oct2016, format(%tdnn/dd/yy)) ///
    ytitle("Number of cases per week" "(Zika in units of 50 cases)" ) lp(dash solid) color(black black)
```

/*10.0 OBTAINING CROSS-CORRELATIONS BEFORE AND AFTER ACCOUNTING FOR THE EFFECT OF TIME. COLOMBIAN DATA*/

```
frencurv, xvar(week) gen(splzika)  power(3) refpts(5 10 15 20 25 30 40 50 60)
glm zika splzika*, fam(poisson) scale(x2) nocons
predict zikaxb, xb
gen zikaxbnumber =exp(zikaxb)
gen zikares=zika-zikaxbnumber
tsline zika zikaxbnumber
frencurv, xvar(week) gen(splmicro)  power(3) refpts(20 25 30 40 50 60)
glm micro splmicro3-splmicro8, fam(poisson) scale(x2) nocons
predict microxb, xb
gen microxbnumber =exp(microxb)
gen microres=micro-microxbnumber
tsline micro microxbnumber
expand 2, gen(expanded)
label var zika "Zika infections"
label var micro "microcephaly"
xcorr zika micro, lags(38) saving(zika_micro.gph, replace) ti("Observed cases") ///
                        ytitle("Cross-correlation of Zika virus infections""and newborn microcephaly") ///
                        ylabel("", axis(2))  xtitle("Lag in weeks") msize(small) ///
                        xcorr zikares microres, lags(38) saving(zikares_microres.gph, replace) ///
                        ti("Residual cases") ytitle("") ylabel("", axis(2)) xtitle("Lag in weeks") msize(small) ///
                        gr combine zika_micro.gph zikares_microres.gph, ///
                        saving(Appendix_Figure3_JPHE.gph, replace)  ///
                         ti("Appendix Figure 3. Cross-correlograms of weekly observed and residual"  ///
                         "number of cases of Zika virus infections and newborn microcephaly in"       ///
                         "Colombia (2015-2016).", justification(left) size(medsmall) color(black))
drop if expanded==1
```

/*11.0 FITTING AUTOREGRESSIVE POISSON MODEL FOR THE LAG TIME WITH THE HIGHEST CROSSCORRELATION. COLOMBIAN DATA*/

```
capture program drop maxeffects
qui program define maxeffects
use t1_t2.dta, clear
qui for num 0/68: qui gen microlagX = micro[_n-X]
qui for num 0/68: qui gen DmicrolagX = D.microlagX
qui for num 0/68: qui gen zika50lagX = zika50[_n-X]
qui for num 0/68: qui gen Dzika50lagX = D.zika50lagX
qui arpois micro Dzika50lag19, prev delete ar(1)
```

```
qui lincom X1
scalar rr=exp(r(estimate))
scalar lbrr=exp(r(estimate)-1.96*r(se))
scalar ubrr=exp(r(estimate)+1.96*r(se))
dis as text _col(2) "Rate ratio of microcephaly per change of 50 cases in Zika 19 weeks before: "  as res _col(80)
`lag' _col(15) %6.4f rr " (" %6.4f lbrr ", " %6.4f ubrr ")"
end
maxeffects
```

/*12.0 EVALUATING THE IMPACT OF ALTITUD ON THE PREVALENCE OF MICROCEPHALY IN COLOMBIA*/
/*12.1.Creates a dataset with the latitud and longitude for every county (municipio) and Department in Colombia. Excel files available from the author upon request.*/

```
clear    all
import excel using DIVIPOLA_20170330_NEW.xlsx, first /*From DANE; modified to eliminate un-needed rows:
http://geoportal.dane.gov.co:8084/Divipola/*/
destring Latitud, gen(latitude) dpcomma
destring Longitud, gen(longitude) dpcomma
collapse longitud latitud, by( CodigoMunicipio NombreDepartamento NombreMunicipio)
gen obs=_n
save elevdata1, replace
```

```
/*NOTE: RUNNING CODE BELOW REQUIRES A GOOGLE MAPS KEY AND PAYMENT*/
/*12.2 Obtaining the elevation of each county from Google maps*/
clear all
use elevdata1, clear
capture shell erase temp*.dta
goelevation1, lat(latitude) lng(longitude) apikey("YOUR API KEY HERE")saving(temp1)
use temp1, clear
gen obs=seno
sort obs
merge 1:1 obs using elevdata1
assert _merge==3
drop _merge
save elevdata2, replace
```

```
/*12.3 Creating a datase with data on population in each county in Colombia*/
clear
import excel Municipal_area_1985-2020_New.xlsx, first /*From DANE; modified to eliminate un-needed rows:
http://geoportal.dane.gov.co:8084/Divipola/*/
rename AJ y2016
rename AK y2017
keep DP DPNOM DPMP MPIO y2016 y2017
rename DPMP codmun
sort codmun
drop if y2016==.

destring DP, gen(codedep)
gen deppop2017=.
levelsof codedep, loc(codedep)
```

```
foreach i of loc codedep{
qui sum y2017 if codedep==`i'
replace deppop2017=r(sum) if codedep==`i'
}
save pop1, replace


/*12.4 Generating a variable with the proportion of population living above 2000 meters in each
Department*/
use elevdata2, clear
rename CodigoMunicipio codmun
sort codmun
merge codmun using pop1
assert _merge==3
drop _merge
distinct codmun
assert r(ndistinct)==r(N)
rename NombreDepartamento departamento
rename NombreMunicipio municipio
destring DP, gen(depcode)
sort departamento
label define depcodelb  91 "Amazonas"  5 "Antioquia" 81 "Arauca"  8 "Atlantico" 11 "Bogota" ///
                        13 "Bolivar" 15 "Boyaca" 17 "Caldas" 18 "Caqueta" 85 "Casanare" 19 "Cauca" ///
                        20 "Cesar" 27 "Choco" 25 "Cordoba" 23 "Cundinamarca" 94 "Guainia" ///
                        95 "Guaviare" 41 "Huila" 44 "LaGuajira"  47 "Magdalena" 50 "Meta"  ///
                        52 "Narino" 54 "NSantander" 86 "Putumayo" 63 "Quindio" 66 "Risaralda" //
                        88 "SAyProviden" 68 "Santander" 70 "Sucre" 73 "Tolima" 76 "ValledeCauca"
vlabel depcode depcodelb
gen highele=0
replace highele=1 if elevation>2000
label var highele "High elevation"
label define ynlb 0 No 1 Yes
vlabel highele ynlb
gen phigh=.
levelsof depcode, loc(depcode)
foreach i of loc depcode {
        qui sum y2016 if depcode==`i' & highele==1
        scalar t`i'_1=r(sum)
        qui sum y2016 if depcode==`i'
        scalar total_`i'=r(sum)
        replace phigh=t`i'_1/(total_`i') if depcode==`i'
        }
sum phigh if depcode==5 /*Check result is correct for Antioquia*/
gen phighcat=2
replace phighcat=1 if phigh>0 & phigh<0.50
replace phighcat=0 if phigh>=0.50
tabulate phighcat, gen(phgroup)
save elevdata3, replace
```

```
/*12.5 Creating dataset with number of cases of microcephaly (Cuevas, Supplemental table) and infant
mortality by Department (source:
https://www.minsalud.gov.co/salud/Paginas/CoberturasdelR%C3%A9gimenSubsidiado.aspx
        File: certificacion por departamentos.
        Source for asegurados by April 2017 : http://www.sispro.gov.co/Pages/Aseguramiento/Reporte-
Aseguramiento.aspx*/

clear
input depcode str28 departamento microcases births infmort asegurados
 91 Amazonas          2       695     38.93   69152
  5 Antioquia         24      58541   13.29   6017670
 81 Arauca            0       3266    12.32   241931
  8 Atlantico         31      30604   17.08   2425438
 11 Bogota            45      77754   13.82   7390496
 13 Bolivar           26      25940   13.61   2046824
 15 Boyaca            5       11993   13.77   1103353
 17 Caldas            4       7476    13.39   863564
 18 Caqueta           9       5290    20.01   378619
 85 Casanare          13      4836    12.89   371794
 19 Cauca             10      14078   18.04   1241698
 20 Cesar             30      15235   20.25   1131759
 27 Choco             3       4072    29.39   420780
 25 Cordoba           33      19902   20.85   1621103
 23 Cundinamarca      15      27938   14.08   2134401
 94 Guainía           0       435     40      42167
 95 Guaviare          1       888     21.13   72498
 41 Huila             31      15282   13.3    1076951
 44 LaGuajira         13      10906   24.6    842937
 47 Magdalena         16      15613   21.78   1292821
 50 Meta              13       11760   15.31   908537
 52 Narino            3       13865   16.46   1425525
 54 NSantander        39      15869   16.77   1356156
 86 Putumayo          2       3241    17.79   302038
 63 Quindio           1       4564    15.86   501096
 66 Risaralda         0       8417    15.47   907412
 88 SAyProviden       2       710     17.22   56137
 68 Santander         20      23364   11.89   2009977
 70 Sucre             20      10791   13.43   944815
 73 Tolima            35      13776   14.69   1206810
 76 ValledeCauca      30      39855   12.17   4236715
 97 Vaupes            0       273     30.19   30881
 99 Vichada           0       773     39.79   75208
end
label var infmort "Infant mortality/1000 in 2007"
save microcases_col, replace

/*12.6 Merging the data file with cases and proportion of population by altitude in each Department*/
/*Note: Different data sets are created and models are fitted to verify results are replicable.*/
use elevdata3, clear
sort depcode
```

```
merge m:1 depcode using microcases_col
keep if _m==3
gen paseg=asegurados/deppop2017
collapse (mean) depcode phigh phgroup1 phgroup2 phgroup3 births microcases infmort asegurados
deppop2017 paseg, by(departamento)
poisson micro phgroup2 phgroup3 infmort paseg, exposure(births) irr  /*Results are similar to those from
logistic regression below*/
gen h=0
replace h=1 if phgroup2==1 | phgroup3==1
label var h "Altitude>2000 m?"
label values h ynlb
gen noncases=births-microcases
gen total=micro+noncases
gen w0 = noncases
rename microcases w1
gen id = _n
reshape long w, i(id) j(y)
logistic y phgroup2 phgroup3  infmort paseg [fw=w]  /*Verifying results are consistent with those from poisson
regression above*/
expand w, gen(expanded)
drop if w==0
logistic y phgroup2 phgroup3 infmort paseg /*Verifying results are consistent with those from Poisson and
logistic regression above*/
drop id w noncases total            /*These variables were needed only for reshaping the data*/
label var y "Microcephaly"
label values y ynolb
label var phigh "Proportion of pop above 2000 m"
label var phgroup1 "phigh>=60%"
label values phgroup1 ynlb
label var phgroup2 "phigh>0 and <60%"
label values phgroup2 ynlb
label var phgroup3 "phigh==0%"
label values phgroup3 ynlb
label var births "Number of births"
label var expanded "Expanded observation?"
label values expanded ynlb
label var paseg "Proportion pop with health insurance"
label var infmort "Infant mortality in Department"
label var asegurados "Number with health insurance"
label var deppop2017 "Deparment population 2017"
label data "Combined data on microcephaly cases and proportion of population living above 2000 meters by
Department"
save elevmicro, replace
```

/*13.0 SENSITIVITY ANALYSIS FOR MISCLASSIFICATION OF EXPOSURE (LIVING ABOVE 2000 M)*/
```
/*Procedure takes a lot of time. Sampling 1% of non-cases for the purpose of checking pvw runs as expected*/
use elevmicro, clear
sample 1 if y==0
pvw y inf paseg, casesens(1) contsens(1) casespec(1) contspec(1) outcome(y) seed(13323) misclass(h)
othercov(inf paseg)
```

```
/*Models from step above should be identical, since there is no misclassification of exposure*/

use elevmicro, clear
capture program drop san
qui program define san
tempname memhold
tempfile q8
postfile `memhold' double(sens spec or lbor ubor) using `q8'
foreach i in  0.5 0.6 0.7 0.8 0.9 1 {
        foreach j in  0.5 0.6 0.7 0.8 0.9 1 {
        qui pvw y inf paseg, casesens(`i') contsens(`i') casespec(`j') contspec(`j') outcome(y) seed(13323)
misclass(h) othercov(inf paseg)
        qui lincom truecov
        local sens=`i'
        local spec=`j'
        local or=exp(r(estimate))
        local lbor=exp(r(estimate)-(1.96*r(se)))
        local ubor=exp(r(estimate)+(1.96*r(se)))
        //dis as text _col(2) "sens: " `i' " " "spec: " `j' "  OR=" as res %4.2f or " (" %4.2f lbor "," %4.2f ubor ")"
        post `memhold' (`sens') (`spec') (`or') (`lbor')  (`ubor')
        }
}
postclose `memhold'
use `q8', clear
save ansen, replace
end
san

/*13.1 Plotting estimates from sensitivity analysis*/
use ansen, clear
label var sens "Sensitivity"
label var spec "Specificity"
label var or "Odds ratio"

twoway (line or spec if sens==0.6 & spec>0.5 & spec<1) (line or spec if sens==0.7 & spec>0.5 & spec<1)  ///
    (line or spec if sens==0.8 & spec>0.5 & spec<1) (line or spec if sens==0.9 & spec>0.5 & spec<1), ///
    legend( order(1 "Sensitivity=0.6" 2 "Sensitivity=0.7"  3 "Sensitivity=0.8"  4 "Sensitivity=0.9")) ///
    ti("Appendix Figure 4. Odds ratios for altitude and microcephaly corrected for" ///
    "changing levels of misclassification of the exposure. Colombia, 2015-2016.", ///
    justification(left) size(medsmall) color(black)) ///
    xtitle("Specificity of altitude classification", size(medsmall))
capture log close
exit
```

References

26. Silveira MF, Matijasevich A, Horta BL, et al. Prevalence of preterm birth according to birth weight group: a systematic review. Rev Saude Publica 2013;47:992-1003.
27. Mitchell M, Digitizer E. A free open-source software to extract data points from a graph image Hosted on SourceForge. Available online: https://sourceforge.net/projects/digitizer/. 2017.
28. Anuranga C, Jayanthan JV. GOELEVATION: Stata module to compute elevation for latitude and longitude from Google. Statistical Software Components. 2014. Available online: https://ideas.repec.org/c/boc/bocode/s457946.html
29. Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. Stat Med 2010;29:2297-309.
30. Peterson AT, Osorio J, Qiao H, et al. Zika Virus, Elevation, and Transmission Risk. PLoS Curr 2016;8. pii: ecurrents.outbreaks.a832cf06c4bf89fb2e15cb29d374f9de.