

Predictibilidad de la propagación espacial y temporal de la epidemia de influenza A H1N1 en la Argentina por el método de percolación

EDUARDO CUESTAS^{1*}, MARIO VILARÓ², PABLO SERRA³

¹Área de Epidemiología Clínica y Bioestadística, Hospital Privado Centro Médico de Córdoba, Cátedra de Estadística Médica, Facultad de Medicina, Universidad Católica de Córdoba; ²Laboratorio de Microbiología, Hospital Privado Centro Médico de Córdoba. Av. Naciones Unidas 346 (5016), Córdoba; ³Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba. Medina Allende s/n, Ciudad Universitaria, (5000) Córdoba, Argentina.

*Correspondencia. E-mail: ecuestas@hospitalprivadosa.com.ar

RESUMEN

La epidemia de influenza A H1N1 se expandió rápidamente a nivel mundial dadas las actuales condiciones de alta interconectividad y velocidad de los transportes, imperantes tanto entre las personas como entre los países y las regiones. La diseminación espacial de la epidemia puede ser explicada mediante la teoría de la percolación, que permite estimar un umbral más allá del cual se produce el traspaso de la epidemia entre distintas regiones geográficas. El objetivo de este trabajo fue probar la capacidad predictiva del modelo de percolación aplicado al análisis de la epidemia de influenza A H1N1 registrada en la Argentina en 2009, de acuerdo a los datos relevados por el Ministerio de Salud Pública de la Nación. Para aplicar el mencionado modelo se consideró al país como un conjunto de figuras geométricas irregulares, contiguas y continuas, que pueden representarse en dos dimensiones en una carta geográfica plana. Se analizó la proporción de provincias infectadas en el momento de la percolación con respecto al tiempo y se compararon los valores observados con los esperados mediante ecuaciones de estimación curvilínea en un modelo logístico. La percolación ocurrió en el día 45. El valor esperado que generó el modelo fue de 42,4 días, intervalo de confianza de 95 % 28,5-56,3. La diferencia entre el valor observado y el esperado arrojó un valor de $p = 0,997$. Se concluye que el modelo posee un buen ajuste y una adecuada capacidad predictiva.

Palabras clave: propagación; epidemia; influenza A H1N1, Argentina

ABSTRACT

Spatial and temporal spread predictability of influenza A H1N1 epidemic in Argentina by the percolation method.

The influenza A H1N1 epidemic has spread rapidly worldwide on account of the current conditions of high interconnectivity and transport speed both among people and countries. The spatial spread of the epidemics can be explained by the percolation theory which allows to estimate a threshold beyond which the transmission of the infection among different geographic regions occurs. The aim of this study was to test the predictive ability of the percolation model of influenza A H1N1 epidemic in Argentina according to data gathered by the National Department of Public Health. In the model, the country was considered as a set of irregular, contiguous and continuous geometric figures, which can be represented in two dimensions on a plane. We analyzed the proportion of infected provinces at the moment of percolation in relation to time in days and compared observed and expected values by curvilinear equations in a logistic model. Percolation occurred on day 45. The expected value generated by the model was 42.4 days, 95 % CI 28.5 to 56.3. The difference between observed and expected values was $p = 0.997$. We conclude that the model has good fit and predictive capacity.

Key words: Spread; epidemic; influenza A H1N1, Argentina

INTRODUCCIÓN

El 29 de abril de 2009, la Organización Mundial de la Salud (OMS) anunció la posibilidad de una rápida diseminación global de una cepa del virus de influenza A H1N1 que había sido detectada la semana anterior, por lo que se estableció un nivel de alerta fase 5 (18). La fase 5 indica la transmisión continua de humano a humano de una nueva cepa de influenza de origen animal en una de las regiones de la OMS, tras haberse detectado casos exportados a otra u otras regiones.

El virus de la influenza es altamente mutagénico y presenta un número reproductivo básico (R_0) cercano a 2, considerado moderado, aunque suficiente para propagar rápidamente la infección y comprometer a una gran cantidad de individuos de una población (5).

En la Argentina, la epidemia desbordó la capacidad del sistema sanitario nacional y puso en evidencia la falta de elementos objetivos que pudiesen estimar la capacidad de propagación epidémica en el espacio y en el tiempo. Esta circunstancia limitó la implementación de medidas tales como organizar y establecer, con la anticipación

debida, las ubicaciones geográficas de los centros de referencia diagnóstica y de procesamiento de muestras virológicas, entre muchas otras que hubieran permitido dar una respuesta asistencial más eficiente.

Analizar retrospectivamente el comportamiento de una epidemia aporta información útil ante posibles recurrencias y sirve para el ajuste de nuevas metodologías que faciliten interpretaciones prospectivas. El estudio de la dinámica de la transmisión microbiana tratando de ajustarla a modelos matemáticos ha sido objeto de varios estudios en los últimos años, y aunque los resultados obtenidos son prometedores, la asociación entre la matemática y la microbiología continúa siendo un tema controvertido y complejo. No obstante, la interpretación de los fenómenos epidemiológicos usando patrones matemáticos es un campo en permanente desarrollo (7). En los últimos años se han incorporado a la epidemiología conceptos teóricos provenientes de la física y la matemática para el análisis de las dinámicas asociadas con fenómenos complejos, entre ellos los de caos dinámico, fractales, sistemas disipativos y autoorganización (14).

Un proceso estocástico es un concepto estadístico que sirve para caracterizar una sucesión de variables aleatorias que evolucionan en función de otra variable, que generalmente es el tiempo. Para explicar la propagación de una enfermedad en el espacio y en el tiempo, se ha propuesto modelar la transmisión de las enfermedades infecciosas mediante la dinámica de redes estocásticas utilizando el concepto de conectividad espacial, de acuerdo a la teoría de la percolación aplicada a la epidemiología o geometría del contagio (6).

La percolación es una teoría con la cual es posible estudiar los patrones que exhibe la dinámica de un sistema interconectado al azar (12). Si consideramos una malla de geometría rectangular, cada retícula puede estar "ocupada" o "desocupada" respecto de un factor determinado (en nuestro caso, la presencia o ausencia de infección), y este suceso se puede representar con una probabilidad p . Entonces, considerando a una retícula de ubicación cualquiera en la malla, ocupada con una probabilidad p , y a sus vecinas, desocupadas aleatoria e independientemente con una probabilidad $1-p$, sería posible modelar el comportamiento aleatorio de ocupación de las retículas mediante un algoritmo de recorrido (o propagación de la infección). Dependiendo del tipo de geometría de la malla, el sistema evoluciona en el tiempo hasta que en un momento dado es posible pasar de agregaciones aisladas de retículas ocupadas a conglomerados que atraviesan la malla de un extremo a otro, tanto en sentido horizontal como vertical, evento que se define como umbral de percolación. El valor de p en el que se llega al umbral de percolación se denomina probabilidad crítica (p_c) (9).

Si suponemos entonces que una provincia infectada puede contagiar a las vecinas que establecen continuidad con ella, se establece un sistema de propagación de la

infección. La percolación es un fenómeno de umbral, y para el caso particular de retículas finitas, el valor del umbral de percolación es $p_c = 0,5927$, tal como lo demuestra Turner en un modelo estocástico (16). Esto significa que el 59 % de las provincias deberían estar infectadas para que pueda producirse la transmisión de la infección de norte a sur y de este a oeste.

Se ha propuesto que la percolación puede describir adecuadamente el avance geográfico de la epidemia de influenza A H1N1. Luego de analizar la propagación de la epidemia en los Estados Unidos de América sobre la base de este modelo y considerando a los estados de ese país como celdas, Canals (1) propuso que el modelo de percolación es útil para describir adecuadamente el progreso que tuvo la epidemia, tras demostrar que el umbral de percolación observado coincide con el esperado cuando se ajusta una regresión entre la proporción de estados infectados y el tiempo con el método de probito. Asimismo y basándose en la utilización de países como unidades geográficas, Canals y Canals (3) calcularon, al inicio de la epidemia, el tiempo estimado de percolación mundial, con mucho mejor ajuste que el obtenido mediante el modelo de celdas.

El objetivo de este trabajo fue probar la capacidad predictiva del modelo de percolación aplicado al análisis de la epidemia de influenza A H1N1 registrada en la Argentina en 2009.

MATERIALES Y MÉTODOS

Durante el curso de la epidemia, se registraron las provincias con casos confirmados por el Instituto Nacional de Microbiología en función del tiempo; en dicha situación, la provincia podía cambiar del estado de "no infectada" al de "infectada". Se catalogó a una provincia como "infectada" desde el momento en que aparecía el primer caso de influenza confirmado. Las provincias que fueron cambiando de estado, ante la aparición de casos confirmados, se registraron en orden sucesivo sobre el mapa político de la República Argentina.

Los datos fueron obtenidos de los partes periódicos emitidos por el Ministerio de Salud de la República Argentina (13) durante el transcurso de la epidemia, desde el 29 de abril hasta el 30 de diciembre de 2009. La proporción de provincias infectadas en función del tiempo se calculó sobre un mapa del Sistema de Información Geográfica diseñado por el Instituto Geográfico Nacional (8), vigente a la fecha de la epidemia. Para realizar dicho cálculo, a los 24 distritos se les debe sustraer la provincia de Tierra del Fuego, Antártida e Islas del Atlántico Sur, ya que por su condición insular no contigua no cumple con uno de los postulados básicos de la percolación (24 distritos = 23 provincias + Ciudad Autónoma de Buenos Aires – Tierra del Fuego, Antártida e Islas del Atlántico Sur = 23).

Como definición operativa, se consideró como percolación un estado en el cual se tiene continuidad geográfica de las provincias infectadas; es decir, se puede atravesar el mapa del territorio nacional de un extremo a otro, en los dos sentidos (horizontal y vertical, sea cual fuere el recorrido). El umbral de percolación es, en cambio, la proporción media de provincias infectadas en el límite mínimo en que el fenómeno de percolación comienza (cambia el estado del sistema de no percolante a percolante). Es dable notar que por percolación denominamos a un estado arbitrario de provincias infectadas, mientras que el umbral de

percolación es una media estadística, la cual calculamos por el método de Montecarlo (ver más adelante), y esta cantidad tiene un error estadístico (excepto cuando el sistema es infinito). Finalmente, el tiempo de percolación es el momento temporal en el cual se alcanza la percolación (estado en el que es posible realizar recorridos percolantes de un extremo a otro del sistema).

Se realizaron regresiones lineales entre estas proporciones desde el inicio de la epidemia mediante estimaciones curvilíneas bajo un modelo logístico: $[p = a \cdot t + b]$, donde "p" es la proporción de provincias infectadas, "t" es el tiempo en días de la progresión de la epidemia, "a" es la pendiente de la regresión y "b" es la intercepción en el origen. Cabe aclarar que los modelos de regresión lineal se utilizan para predecir el valor de una característica a partir de otra; en el caso particular que analizamos, la proporción de provincias afectadas sirve para predecir el tiempo de propagación de la infección. Para ello se realiza un cálculo mediante una ecuación lineal, en la cual la pendiente de la línea y su intercepción con los ejes cartesianos ortogonales permiten, al realizar la gráfica de la función, determinar el valor esperado del tiempo de propagación en relación con la proporción de provincias infectadas. A partir de estos se calculan los "estadísticos", que son el coeficiente de regresión (R), el coeficiente de determinación (R^2) y la prueba F, que da la significación estadística del modelo.



Figura 1. Provincias infectadas (en negro) al día 45 de la epidemia. En este momento es posible establecer una continuidad de recorrido, atravesando el mapa de norte a sur y de este a oeste.

El umbral de percolación para la República Argentina se calculó realizando sorteos aleatorios, utilizando un modelo de simulación tipo Montecarlo para percolaciones verticales (norte-sur) y horizontales (este-oeste), y observando que las provincias no se superpusieran, hasta obtener un valor constante en tres dígitos. Luego de aproximadamente 20 corridas, la proporción de éstas sobre el total se estabilizó en la cifra $pc = 0,5927$ (3, 8, 10, 13). Reemplazando este valor en las regresiones, se determinó el tiempo predicho para alcanzar ese umbral como tiempo de percolación (t_p). Para esta predicción se estimó el intervalo de confianza del 95 % según la siguiente fórmula: $IC95\% = \text{tiempo de percolación} \pm 1,96$ multiplicado por la raíz cuadrada del tiempo ($t_p \pm 1,96 \times \sqrt{t}$). Este procedimiento se realizó usando la información correspondiente a los días 10, 15, 20, 25, 30 y 50 desde el inicio de la epidemia, con el fin de estudiar la capacidad predictiva del modelo. Los valores esperados se compararon con los observados mediante la prueba de bondad de ajuste de Hosmer-Lameshow. Se estableció un valor de significación de $p < 0,05$. Para realizar los cálculos, se utilizó el programa de estadística SPSS (SPSS Inc., Chicago, IL, EE.UU.).

RESULTADOS

La percolación nacional ocurrió el día 45 después del primer caso confirmado (7 de mayo) y a los 51 días se produjo la extensión de la epidemia a todo el territorio argentino. En el día de percolación habían cambiado de estado el 45 % de las provincias, con una diferencia porcentual de 14 % respecto del valor esperado (59 %). La proporción de provincias infectadas en esa fecha arrojó entonces un valor observado muy cercano al umbral de percolación esperado, de modo que no hubo una diferencia estadísticamente significativa entre estos valores ($p = 0,997$) (Figura 1).

En el modelo general, la proporción de provincias tuvo un ajuste adecuado, aunque presentó una distorsión en los primeros días, cuando los valores observados cayeron por debajo de los valores predichos, y en los últimos días, cuando los valores observados se encontraron por encima de los esperados (Figura 2).

Los valores predichos para la percolación, calculados con el modelo, fueron razonablemente adecuados cuando se realizaron las predicciones a los 10, 15, 20, 25, 30 y 50 días (Tabla 1). Para estos tiempos, los valores predi-

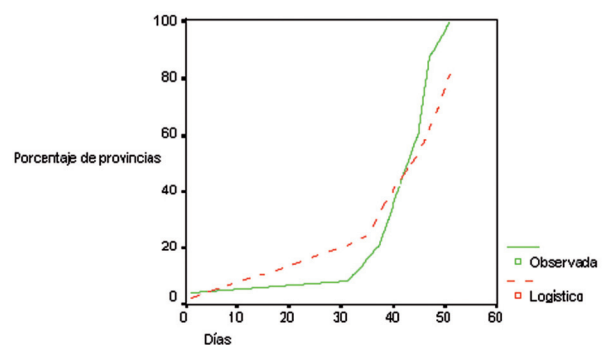


Figura 2. Regresión logística del porcentaje de provincias infectadas en función del tiempo, en días.

chos se ajustaron a los valores observados, por lo que la capacidad predictiva de la proporción de provincias fue estable desde el inicio.

DISCUSIÓN

En este estudio se aplicó por primera vez el método de percolación para determinar el tiempo umbral de la difusión geográfica de la epidemia de influenza A H1N1 que tuvo lugar en la Argentina durante el año 2009.

Con este mismo enfoque, otros autores han realizado adecuadas predicciones del tiempo de percolación, tanto para los EE.UU. (1) como a nivel internacional (3), de forma precisa y estable desde el inicio de la epidemia, especialmente cuando se consideró a los países como unidades geográficas, ya que el modelo basado en celdas de forma regular y simétrica usado en matemática dista de la geometría variable, asimétrica e irregular de los países. En estos trabajos, el cociente entre el tiempo de percolación calculado y el observado estuvo ligeramente por debajo de la unidad a los 10 días, cercano a ésta a los 25 días y algo por encima a los 30 días. El método logarítmico dio previsiones más precisas que las proporcionadas por los modelos basados en los números crudos (sin transformación logarítmica) en el análisis de las variables.

Estos hallazgos coinciden con nuestros resultados. Por esto, debe observarse que las predicciones del tiempo de percolación fueron siempre muy estables y cercanas a la unidad. Este hecho permite deducir que el modelo propuesto por Canals y Canals (3) puede ser mejorado para evitar las distorsiones encontradas durante los primeros días, lo que evidentemente optimizaría la capacidad predictiva del modelo previamente planteado (2).

Una debilidad de nuestro trabajo, al igual que lo observado en otras investigaciones (3), es que se calculó el umbral de percolación para retículas cuadradas ($p_c =$

0,5927), y este fue aplicado de manera indistinta en la estimación del tiempo de percolación, tanto por celdas como por países o provincias. Esta forma de cálculo podría condicionar el resultado de la variable dependiente (tiempo) y distorsionar la capacidad predictiva del modelo. Para evitar este inconveniente, teóricamente cabría estimar el umbral de percolación sobre modelos geográficos de redes libres de escala (17), con un umbral de percolación de 0,3913. Tales modelos parecen representar en forma más realista los sistemas territoriales concretos (15), aunque lamentablemente, por el momento son muy difíciles de llevar a la práctica dada la gran variabilidad de las formas geográficas, mucho más complejas que el simple modelo reticular de cuadrículas (4).

La utilización de este modelo incluye una serie de elementos que aumentan su fortaleza, como la ausencia de suposiciones sobre la distribución de parámetros, sobre las distancias y recorridos y, particularmente, sobre las interrelaciones espacio-temporales. Estos factores podrían claramente explicar el mejor ajuste que presenta el modelo estocástico de percolación con respecto a los modelos epidemiológicos clásicos, que usan ecuaciones diferenciales (modelos SIER: susceptibles, infectados, enfermos, recuperados) para estimar el número de casos en un tiempo determinado, como lo demuestra Canals en un trabajo reciente (2).

Este trabajo presenta una serie de limitaciones que, sin duda, influyen sobre los resultados obtenidos y deben ser mencionadas, como las variables externas concomitantes de tipo meteorológico y demográfico, las distintas prácticas de atención y las diferentes condiciones socioeconómicas, de accesibilidad, de vías de comunicación y de barreras geográficas, cuya complejidad de análisis hace que sean imposibles de incluir en el modelo estudiado. Además, los datos se basan únicamente en casos confirmados, por lo que seguramente se subestiman los casos reales que pueden haberse presentado en las distintas

Tabla 1. Capacidad predictiva del tiempo de percolación (t_p) y su respectivo IC 95 % para las regresiones entre el porcentaje umbral estimado de provincias infectadas y el tiempo, considerando diferentes números de días desde el comienzo de la epidemia hasta su propagación a todo el territorio nacional.

Días	Pendiente	Intercepción	t_p	IC 95 %	R^2	F	p
10	0,9766	0,2436	40,7	34,5-46,9	1,000	-	-
15	0,9633	0,2633	41,3	33,7-48,8	0,800	04,22	0,288
20	0,9474	0,3200	41,9	33,1-50,6	0,787	07,40	0,113
25	0,9406	0,3504	42,2	32,4-52,0	0,810	12,80	0,037
30	0,9365	0,3728	42,3	31,6-53,0	0,825	18,88	0,012
50	0,9344	0,3883	42,4	28,5-56,3	0,848	27,92	0,003

t_p : Tiempo de percolación, IC 95 %: Intervalo de confianza de 95 %, R^2 : Coeficiente de determinación, F: Valor de la distribución de probabilidad F para la prueba de Hosmer-Lameshow, p : Probabilidad de significación estadística (o probabilidad de cometer un error tipo I en el modelo predictivo), valor observado/valor esperado.

regiones geográficas, con lo cual la onda de propagación registrada podría no ajustarse necesariamente a la propagación real de la epidemia en el tiempo y el espacio.

CONCLUSIONES

En conclusión, el método aquí descrito, que hace uso extensivo de los datos de la epidemia observada para estimar la dinámica temporal y geográfica del proceso de difusión, constituye un nuevo enfoque para la previsión de la difusión de las epidemias de gripe. El cálculo de la percolación sería probablemente aplicable a la predicción de otras enfermedades transmisibles. Al igual que cualquier otro método de predicción, la percolación requiere de una recolección en tiempo real de datos para hacer las predicciones a futuro. Por último, el método propuesto debe ser contrastado con futuras investigaciones que evalúen las interacciones espacio-temporales con el menor sesgo posible (11).

BIBLIOGRAFÍA

1. Canals M. Inicio de la pandemia AH1N1: Álgebra, cálculo y geometría del contagio. *Rev Med Chile* 2009; 137: 852-6.
2. Canals M. Predictibilidad a corto plazo de casos de influenza AH1N1 basado en modelos determinísticos. *Rev Chil Infectol* 2010; 27: 114-20.
3. Canals M, Canals A. Percolación de la epidemia de influenza AH1N en el mundo: Utilidad de los modelos predictivos basados en conectividad espacial. *Rev Med Chile* 2010; 138: 573-80.
4. Eguíluz V, Klemm K. Epidemic threshold in structured scale-free networks *Phys Rev Lett* 2002; 89: 108701. [<http://prl.aps.org/pdf/PRL/v89/i10/e108701>].
5. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Chapela IB, Zavala EP, Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C. WHO Rapid pandemic assessment collaboration. Pandemic potential of a strain of Influenza A (H1N1): early findings. *Science* 2009; 324: 1557-61.
6. Grassberger P. On the critical behavior of the general epidemic process and dimensional percolation. *Math Biosci* 1983; 63: 157-72.
7. Grassly N, Fraser C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol* 2008; 6: 477-87.
8. Instituto Geográfico Nacional. República Argentina. <http://www.ign.gov.ar>. Consultado el 11 de noviembre de 2010.
9. Kenah E, Robins JM. Second look at spread of epidemics on networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007; 76:036113. [Online] <http://pre.aps.org/pdf/PRE/v76/i3/e036113>.
10. Kesten H. *Percolation theory for mathematicians*. 1st edition. Boston, Birkhauser, 1982, p. 113-28.
11. Kuldorff M, Hjalmarsson U. The Knox model and other test for space-time interaction. *Biometrics* 1999; 55: 544-52.
12. Meyers LA. Contact network epidemiology: bond percolation applied to infectious disease prediction and control. *Bull Am Math Soc* 2007; 44: 63-86.
13. Ministerio de Salud de la Nación. República Argentina. <http://www.msal.gov.ar>. Consultado el 11 de noviembre de 2010.
14. Morin E. *Introducción al pensamiento complejo*. 6^a edición. Barcelona, Ghedisa Ed, 2008, p. 25-50.
15. Serrano A, Boguñá M. Percolation and epidemic thresholds in clustered networks. *Phys Rev Lett* 2006; 97:088701-1-4. [Online] <http://prl.aps.org/pdf/PRL/v97/i8/e088701>.
16. Turner MG. Spatial and temporal analysis of landscape patterns. *Landsc Ecol* 1990; 4: 21-30.
17. Warren CP, Sander LM, Sokolov IM. Geography in scale-free networks model. *Phys. Rev.* 2002; E 66, 056105.
18. World Health Organization. 2009, en <http://www.who.int/esr/disease/swineflu/>. Consultado el 11 de noviembre de 2010.