



Original

Sistema de ayuda a la toma de decisiones sanitarias. Propuesta de umbrales de riesgo epidemiológico ante SARS-CoV-2



María Isolina Santiago Pérez^a, Esther López-Vizcaíno^b, Alberto Ruano-Ravina^{c,d,e,*} y Mónica Pérez-Ríos^{c,d,e}

^a Servicio de Epidemiología, Dirección General de Salud Pública, Consellería de Sanidade, Xunta de Galicia, Santiago de Compostela, España

^b Servicio de Difusión e Información, Instituto Galego de Estadística, Xunta de Galicia, Santiago de Compostela, España

^c Área de Medicina Preventiva y Salud Pública, Universidad de Santiago de Compostela, Santiago de Compostela, España

^d CIBER de Epidemiología y Salud Pública, CIBERESP, Madrid, España

^e Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 16 de noviembre de 2020

Aceptado el 29 de diciembre de 2020

On-line el 23 de enero de 2021

Palabras clave:

Covid 19
España
Modelización
Niveles de alerta
Tasas de incidencia

R E S U M E N

Introducción: La pandemia por SARS-CoV-2 es el mayor desafío sanitario en los últimos 100 años, ocasionando el mayor exceso de mortalidad no bélica en este período en el mundo occidental. Ante una enfermedad de elevada contagiosidad y asintomática en un tercio de los casos, es fundamental disponer de modelos que predigan su evolución. Pretendemos desarrollar un modelo de predicción de infección por COVID-19 en provincias españolas.

Método: Análisis de componentes principales funcional de datos epidemiológicos de las provincias españolas en función de su curva epidémica entre el 24 de febrero y el 8 de junio. Con este método se han clasificado las provincias en función de su evolución (peor, intermedia y mejor). Se han empleado los datos del Centro Nacional de Epidemiología.

Resultados: Se identificaron 2 componentes que explican el 99% de la variabilidad de las 52 curvas. La primera componente es la tendencia global de la tasa de incidencia, y la segunda componente es la velocidad de crecimiento o decrecimiento de la incidencia durante el período. Se identificaron 10 provincias en el grupo de peor evolución y 17 en el de evolución intermedia. Los valores umbrales de la tasa de incidencia a 7 días fueron 134 casos/100.000 habitantes para un nivel de alerta 1 (medio) y 167 para el nivel 2 (alto), consiguiendo un elevado poder de discriminación entre provincias.

Conclusiones: Estos niveles de alerta podrían ser de utilidad para decidir medidas que puedan afectar a la movilidad de la población, siempre y cuando haya una situación de transmisión comunitaria de SARS-CoV-2. Esta información sería intercomparable entre áreas sanitarias o comunidades autónomas.

© 2021 SEPAR. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

A Proposed Epidemiologic Risk Threshold for SARS-CoV-2 for Assisting Healthcare Decision-Making

A B S T R A C T

Introduction: The SARS-CoV-2 pandemic is the most important health challenge observed in 100 years, and since its emergence has generated the highest excess of non-war-related deaths in the western world. Since this disease is highly contagious and 33% of cases are asymptomatic, it is crucial to develop methods to predict its course. We developed a predictive model for Covid-19 infection in Spanish provinces.

Methods: We applied main components analysis to epidemiological data for Spanish provinces obtained from the National Centre of Epidemiology, based on the epidemiological curve between 24 February and 8 June 2020. Using this method, we classified provinces according to their epidemiological progress (worst, intermediate, and good).

Keywords:

Covid-19
Spain
Modelization
Alert levels
Incidence rates

* Autor para correspondencia.

Correo electrónico: alberto.ruano@usc.es (A. Ruano-Ravina).

Results: We identified 2 components that explained 99% of variability in the 52 epidemiological curves. The first component can be interpreted as the crude incidence rate trend and the second component as the speed of increase or decrease in the incidence rate during the period analysed. We identified 10 provinces in the group with the worst progress and 17 in the intermediate group. The threshold values for the 7-day incidence rate for an alert 1 (intermediate) were 134 cases/100,000 inhabitants, and 167 for alert 2 (high), respectively, showing a high discriminative power between provinces.

Conclusions: These alert levels might be useful for deciding which measures may affect population mobility and other public health decisions when considering community transmission of SARS-CoV-2 in a given geographical area. This information would also facilitate intercomparison between healthcare areas and Autonomous Communities.

© 2021 SEPAR. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La pandemia por SARS-CoV-2 ha ocasionado un alto impacto en la salud pública y en la economía a nivel mundial y su trascendencia futura probablemente marque una generación. El 11 de marzo de 2020 la OMS declara la enfermedad producida por el SARS-CoV-2 como pandemia¹, mientras en los países europeos ya existía transmisión descontrolada del virus.

Tras un fallo colosal de los mecanismos de detección, alarma y control de la enfermedad, agudizado por la falta de pruebas diagnósticas, comenzó la transmisión comunitaria¹ en la mayor parte de los países, lo que obligó a tomar medidas excepcionales de salud pública, como el confinamiento forzoso de la mayor parte de la población para cortar las cadenas de transmisión del SARS-CoV-2². En España, el confinamiento se aplicó a nivel nacional, pero con una situación epidemiológica muy desigual entre las provincias españolas³. En algunas provincias este confinamiento llegó tarde y no se pudo evitar el colapso del sistema sanitario⁴, mientras que, en otras, debido a la baja transmisión comunitaria, el SARS-CoV-2 causó un impacto en hospitalizaciones y fallecimientos mucho menor⁵. El confinamiento logró progresivamente el control de la enfermedad, pero también un indudable impacto económico⁶.

Disponer de umbrales epidemiológicos de fácil interpretación y que, con pocos datos, permitan predecir la evolución de la pandemia en áreas pequeñas (a nivel de provincia o incluso de áreas más pequeñas), podría permitir a las autoridades sanitarias actuar de forma más eficiente a través del establecimiento de medidas de salud pública más o menos drásticas ante aumentos de incidencia que conlleven transmisión comunitaria no controlada. Hay que destacar que estos umbrales reflejan múltiples características intrínsecas de un país y de una población, desde los comportamientos y contactos sociales hasta la propia estructura de la atención primaria y atención hospitalaria, por lo que umbrales epidemiológicos «de disparo» en un país concreto, no tienen necesariamente por qué ser aplicables en nuestro país. Por esta razón, desarrollar umbrales o niveles de alerta propios podría evitar que las autoridades sanitarias tomaran decisiones erróneas como, por un lado, confinamientos innecesarios (falso positivo) y, por otro, decidir no confinar cuando esta acción es necesaria (falso negativo).

El objetivo de este trabajo es clasificar las provincias españolas en función de la evolución de la pandemia de SARS-CoV-2 durante el período de febrero a junio de 2020 y proponer valores umbrales en las tasas de incidencia que indiquen diferentes niveles de alerta para la vigilancia epidemiológica, que pueden ser útiles para la toma de decisiones que afecten a la movilidad de la población u otras medidas.

Métodos

Fuentes de información

La base de casos diarios de COVID-19 por provincias y ciudades autónomas (en adelante se denominan conjuntamente provincias)

se descargó el 1 de julio de 2020 de la página web del Centro Nacional de Epidemiología (CNE) y contenía información desde el 31 de enero hasta el 28 de junio de 2020³. Estos datos proceden de la encuesta epidemiológica de caso que cada comunidad autónoma (CA) cumplimenta ante la identificación de un caso de COVID-19 y que declaran a la Red Nacional de Vigilancia Epidemiológica (RENAVE) que gestiona el CNE. En la base se incluyen, para cada día y provincia: el número de casos totales (confirmados o probables); el número de casos con prueba de laboratorio PCR o técnicas moleculares; el número de casos con prueba de laboratorio de test rápido de anticuerpos; el número de casos con otras pruebas de laboratorio, mayoritariamente por detección de antígeno o técnica Elisa; y el número de casos sin información sobre la prueba de laboratorio. Para este trabajo se utilizó el número total de casos excluyendo los diagnosticados con test rápidos debido a su limitada fiabilidad con fines diagnósticos ya que no aportaban información sobre si la enfermedad estaba activa. La fecha de declaración de los casos es la de inicio de síntomas o, en su defecto, la fecha de diagnóstico menos 6 días.

Para el cálculo de las tasas se usaron las estimaciones de población residente a 1 de enero de 2020 del Instituto Nacional de Estadística⁷. Las tasas de incidencia a 7 días por fecha de diagnóstico y referidas al día 15 de marzo de 2020 se calcularon a partir de los datos publicados por provincia por el proyecto colaborativo EsCOVID19data⁸. Algunos datos que no estaban disponibles para esa fecha se completaron con la información publicada por las comunidades autónomas (CCAA) en su página web oficial. Las CCAA en las que se buscó información específica fueron: Andalucía, Aragón, Asturias, Cantabria, Castilla y León, Castilla-La Mancha, Cataluña y Comunidad Valenciana.

Las prevalencias de la primera ronda del estudio nacional de seroepidemiología de la infección por SARS-COV-2 en España (ENE-COVID19)⁹ se obtuvieron del informe preliminar del 13 de mayo de 2020. Estas prevalencias son necesarias para calcular la tasa de detección de COVID-19 para cada provincia.

Análisis estadístico

El análisis estadístico de este trabajo comprende 2 procedimientos diferenciados, con datos distintos para cada uno de ellos. En primer lugar, se realizó una clasificación de las provincias españolas en función de la evolución de la pandemia en cada una de ellas, usando como datos la base de casos diarios de la RENAVE. En una segunda etapa, se utilizaron los grupos identificados para calcular puntos de corte en la tasa de incidencia a 7 días que tenían las provincias el día 15 de marzo de 2020; en este caso se consideró como fecha de referencia de los casos la de diagnóstico, y no la de inicio de síntomas, razón por la cual no se pudieron calcular las tasas a partir de la base de casos dia-

rios de la RENAVE. A continuación, se describen con detalle los 2 procedimientos.

La clasificación de las provincias españolas en función de la evolución de la pandemia se realizó a partir de las 52 curvas de incidencia acumulada (casos por 100.000). Siguiendo el criterio del CNE, se consideró como fecha de inicio del análisis el primer día con incidencia acumulada de 5 casos COVID-19/100.000 habitantes en la CA a la que perteneciese cada provincia; el número de días de análisis (92) se calculó como el número de días transcurridos desde la primera fecha con casos (24 de febrero) hasta el 25 de mayo, cuando todas las CCAA estaban al menos en la Fase I de la desescalada. De este modo, la fecha de fin del período de estudio es igual o posterior al 25 de mayo para todas las provincias, y todas tienen el mismo número de días de análisis.

El método de clasificación utilizado comprende 2 pasos. En el primero se extrajeron las componentes principales de las 52 curvas de incidencia acumulada mediante un método funcional de reducción de la dimensión^{10–12}. Cada componente es una combinación lineal de las 52 curvas, y el coeficiente de cada provincia en esa expresión es la coordenada de la provincia asociada a la componente. De este modo, se simplifica o reduce la información de las provincias, que pasa de ser una curva o variable funcional (es decir, una función en el tiempo) a un número pequeño k de coordenadas o, dicho de otra forma, variables unidimensionales.

A partir de los datos de cada provincia se realizó una clasificación en 3 grupos (peor evolución, evolución intermedia y mejor evolución) mediante un análisis clúster de medianas con distancia euclídea¹³.

A continuación, se calculó para cada provincia la tasa de incidencia a 7 días en el día de declaración del estado de alarma (15-03-2020). Teniendo en cuenta el infradiagnóstico de casos en la primera ola de la pandemia, la tasa de cada provincia se corrigió dividiéndola entre la tasa de detección estimada para su CA desde el inicio de la pandemia hasta el 11 de mayo (factor de corrección = $1/\text{tasa de detección}$). La tasa de detección de una CA se estimó como el cociente entre el número de casos diagnosticados hasta el 11 de mayo (informe n.º 103 del Ministerio de Sanidad)¹⁴ y el número de casos estimados a partir de las prevalencias de infección por SARS-CoV-2 en la primera ronda del estudio ENE-COVID19, que terminó el día 11 de mayo⁹.

Las tasas de incidencia provinciales corregidas a 7 días se utilizaron para determinar, mediante la metodología de las curvas ROC, un punto de corte o valor umbral en las tasas que identifique a las provincias de peor evolución de la pandemia (nivel de alerta 2) y a las de evolución intermedia (nivel de alerta 1). El criterio para el cálculo del punto de corte de la curva ROC fue maximizar el índice de Youden (sensibilidad + especificidad - 1). Si la sensibilidad es máxima se minimiza la tasa de falsos negativos, es decir, el porcentaje de provincias con una tasa por debajo del umbral y que realmente tuvieron mala evolución; de igual modo, si la especificidad es máxima se minimiza la tasa de falsos positivos, que en este caso sería el porcentaje de provincias con una tasa por encima del umbral pero que no tuvieron mala evolución. Además, se calcularon medidas de precisión asociadas a los puntos de corte: sensibilidad, especificidad y valores predictivos positivo y negativo, con intervalos de confianza del 95% (IC95%). También se presenta el área bajo la curva ROC con su IC95%. El análisis se realizó con Stata v14.2 y con R (paquetes `fda.usc` y `cluster`).

Resultados

Entre el 31 de enero y el 28 de junio de 2020 se notificaron de forma individual a la RENAVE, en función de la fecha de inicio de síntomas, 269.976 casos totales de COVID-19, de los que 20.670 (7,7%) fueron diagnosticados con pruebas rápidas y se descartaron para

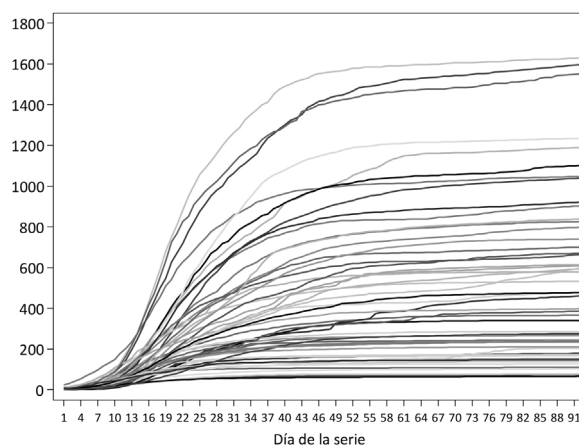


Figura 1. Incidencia acumulada de casos de COVID-19 por 100.000 habitantes en las 52 provincias españolas en un período de 92 días, que varía por provincia entre el 24 de febrero y el 8 de junio.

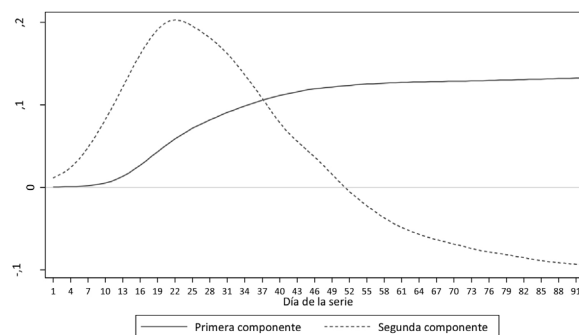


Figura 2. Componentes principales de las 52 curvas de incidencia acumulada de casos de COVID-19 por 100.000 habitantes en un período de 92 días, que varía por provincia entre el 24 de febrero y el 8 de junio.

este análisis. Los 249.306 casos restantes suponen una incidencia acumulada de 526,74 casos por 100.000 habitantes.

Las series de incidencia acumulada para la clasificación de las provincias incluyen 92 días, con fechas de inicio entre el 24 de febrero (Madrid) y el 11 de marzo (Ceuta); en 9 provincias, la fecha de inicio fue el 8 de marzo. En ese período, las provincias con mayor incidencia acumulada fueron Segovia, Ciudad Real y Soria, las 3 con más de 1.500 casos por 100.000, mientras que Huelva, Almería y Las Palmas tuvieron en ese período menos de 100 casos/100.000 (tabla 1 y fig. 1). A nivel estatal, el pico de casos se alcanzó el 20 de marzo, con 10.390; por provincias, la fecha del pico varía entre el 13 de marzo en Melilla y el 21 de abril en Huesca.

Al aplicar el análisis de componentes principales funcional se identificaron 2 componentes (fig. 2) que explican el 99% de la variabilidad de las 52 curvas. La primera componente se puede interpretar como la tendencia global de la incidencia acumulada, y la segunda componente como la velocidad de crecimiento o decrecimiento de la incidencia a lo largo del período de análisis.

Tras aplicar el análisis clúster a las coordenadas de las 2 componentes principales se identificaron 10 provincias en el grupo de peor evolución, 17 en el de evolución intermedia y 25 en el de mejor evolución. Los grupos están definidos fundamentalmente por la coordenada de la primera componente, como se puede observar en la figura 3. En el mapa de España (fig. 4) se observa que las provincias de mejor evolución se concentran en el sureste de la península (Andalucía, Murcia, Comunitat Valenciana, además de Badajoz y Tarragona) y en el noroeste (Galicia – excepto Ourense – Asturias, Zamora, además de Navarra), y las de peor evolución forman un arco

Tabla 1
 Fechas de inicio y fin del período de estudio de 92 días, número acumulado de casos de COVID-19 en ese período e incidencia acumulada por 100.000 habitantes en las 52 provincias. Tasas de incidencia a 7 días el 15 de marzo de 2020, sin corregir y corregidas

Provincia	Período de estudio	Período de estudio		Población	Valores acumulados en todo el período		Tasa de incidencia a 7 días × 10 ⁵ a día 15-03-2020	
		Inicio	Fin		Casos	Casos × 10 ⁵	Sin corregir	Corregida
1	Araba	29/02/2020	30/05/2020	329.857	3.394	1.028,9	93,2	619,1
2	Albacete	01/03/2020	31/05/2020	389.830	4.624	1.186,2	31,9	429,5
3	Alicante	05/03/2020	04/06/2020	1.885.214	3.885	206,1	10,3	120,2
4	Almería	08/03/2020	07/06/2020	715.406	525	73,4	2,9	53,8
5	Ávila	02/03/2020	01/06/2020	158.930	1.746	1.098,6	6,4	61,0
6	Badajoz	07/03/2020	06/06/2020	670.782	907	135,2	3,6	39,0
7	Baleares, Illes	07/03/2020	06/06/2020	1.210.750	2.100	173,4	5,3	78,8
8	Barcelona	02/03/2020	01/06/2020	5.635.043	44.647	792,3	34,5	282,6
9	Burgos	02/03/2020	01/06/2020	355.777	2.141	601,8	33,3	318,8
10	Cáceres	07/03/2020	06/06/2020	390.986	2.052	524,8	16,6	181,2
11	Cádiz	08/03/2020	07/06/2020	1.254.628	1.259	100,3	3,7	68,8
12	Castellón	05/03/2020	04/06/2020	574.900	1.550	269,6	5,5	64,0
13	Ciudad Real	01/03/2020	31/05/2020	494.128	7.872	1.593,1	24,7	331,4
14	Córdoba	08/03/2020	07/06/2020	784.256	1.332	169,8	9,6	178,8
15	Coruña, A	07/03/2020	06/06/2020	1.123.480	4.048	360,3	9,7	59,8
16	Cuenca	01/03/2020	31/05/2020	199.828	1.801	901,3	22,5	301,8
17	Girona	02/03/2020	01/06/2020	767.119	4.533	590,9	17,4	142,5
18	Granada	08/03/2020	07/06/2020	925.059	2.422	261,8	8,8	164,1
19	Guadalajara	01/03/2020	31/05/2020	262.403	1.729	658,9	46,6	626,1
20	Gipuzkoa	29/02/2020	30/05/2020	716.530	2.430	339,1	7,7	51,2
21	Huelva	08/03/2020	07/06/2020	528.059	403	76,3	1,5	28,4
22	Huesca	04/03/2020	03/06/2020	220.657	1.007	456,4	6,7	82,0
23	Jaén	08/03/2020	07/06/2020	628.841	1.425	226,6	7,1	132,9
24	León	02/03/2020	01/06/2020	459.141	2.805	610,9	10,3	98,7
25	Lleida	02/03/2020	01/06/2020	435.607	2.594	595,5	15,1	123,3
26	Rioja, La	27/02/2020	28/05/2020	315.926	3.888	1.230,7	65,7	170,9
27	Lugo	07/03/2020	06/06/2020	328.153	789	240,4	5,8	35,7
28	Madrid	24/02/2020	25/05/2020	6.747.425	69.762	1.033,9	79,5	933,0
29	Málaga	08/03/2020	07/06/2020	1.683.271	2.644	157,1	14,6	271,7
30	Murcia	09/03/2020	08/06/2020	1.504.607	1.591	105,7	4,8	67,1
31	Navarra	02/03/2020	01/06/2020	656.487	5.391	821,2	41,0	307,3
32	Ourense	07/03/2020	06/06/2020	306.802	1.743	568,1	8,2	50,2
33	Asturias	08/03/2020	07/06/2020	1.018.775	2.392	234,8	14,9	116,4
34	Palencia	02/03/2020	01/06/2020	159.846	1.062	664,4	3,1	29,9
35	Palmas, Las	06/03/2020	05/06/2020	1.151.352	743	64,5	3,4	59,7
36	Pontevedra	07/03/2020	06/06/2020	943.809	2.662	282,0	6,9	42,4
37	Salamanca	02/03/2020	01/06/2020	331.048	3.045	919,8	9,4	90,3
38	Sta. Cruz de Tenerife	06/03/2020	05/06/2020	1.085.958	1.582	145,7	7,4	130,9
39	Cantabria	07/03/2020	06/06/2020	582.357	2.296	394,3	7,9	65,5
40	Segovia	02/03/2020	01/06/2020	154.228	2.505	1.624,2	17,6	168,7
41	Sevilla	08/03/2020	07/06/2020	1.957.197	2.430	124,2	4,5	83,1
42	Soria	02/03/2020	01/06/2020	89.912	1.389	1.544,8	5,6	53,9
43	Tarragona	02/03/2020	01/06/2020	814.300	2.239	275,0	4,8	39,1
44	Teruel	04/03/2020	03/06/2020	133.291	629	471,9	11,2	136,1
45	Toledo	01/03/2020	31/05/2020	699.195	4.885	698,7	18,3	246,5
46	Valencia	05/03/2020	04/06/2020	2.568.536	5.976	232,7	10,6	124,3
47	Valladolid	02/03/2020	01/06/2020	520.716	4.349	835,2	7,7	73,6
48	Bizkaia	29/02/2020	30/05/2020	1.142.923	8.418	736,5	11,1	74,0
49	Zamora	02/03/2020	01/06/2020	171.630	657	382,8	4,7	44,9
50	Zaragoza	04/03/2020	03/06/2020	976.498	4.628	473,9	13,6	165,2
51	Ceuta	11/03/2020	10/06/2020	84.032	168	199,9	1,2	9,7
52	Melilla	07/03/2020	06/06/2020	84.496	122	144,4	9,2	123,9

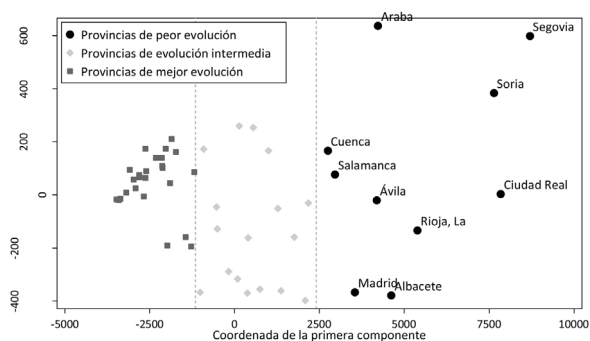


Figura 3. Coordenadas de las 52 provincias en las 2 componentes principales de las curvas de incidencia acumulada y clasificación en 3 grupos a partir de las coordenadas.

desde Salamanca a Álava pasando por Madrid, además de Cuenca, Albacete y Ciudad Real.

En España se diagnosticaron 227.958 casos de COVID-19 desde el inicio de la pandemia hasta el 11 de mayo y se estimó que el número de infectados hasta esa fecha fue de 2.366.499 (prevalencia del 5% del estudio ENE-COVID19), lo que resulta en una tasa de detección del 9,6%. Las tasas de detección por CA varían entre el 5,4% en Andalucía y el 38,2% en La Rioja (tabla 2) y los factores de corrección varían, por tanto, entre 18,6 y 2,6 (tabla 1).

La mediana de las tasas de incidencia corregidas provinciales a 7 días el día 15 de marzo de 2020 fue de 236,4 casos por 100.000 en el grupo de peor evolución, 136,1 en el grupo intermedio y 67,1 en el de mejor evolución, con un recorrido intercuartílico de 339,2; 172,6 y 79,1 respectivamente. Para el cálculo de los valores umbrales en las tasas mediante curvas ROC se consideraron 2 situaciones que

Tabla 2

Prevalencia y número estimado de casos de infección por COVID-19 en España en la primera ronda del estudio ENE-Covid19, número de casos diagnosticados hasta el 11 de mayo, tasa de detección estimada (%) y su inverso (corrección)

Comunidad autónoma	Estudio ENE-Covid19 ronda 1		Hasta el 11-05-2020		
	% infecciones	N.º infecciones	N.º de casos	% detección	Corrección
1 Andalucía	2,7	228.871	12.293	5,4	18,6
2 Aragón	4,9	65.192	5.359	8,2	12,2
3 Asturias	1,8	18.338	2.351	12,8	7,8
4 Islas Baleares	2,4	29.058	1.953	6,7	14,9
5 Canarias	1,8	40.272	2.268	5,6	17,8
6 Cantabria	3,2	18.635	2.246	12,1	8,3
7 Castilla y León	7,2	172.888	18.048	10,4	9,6
8 Castilla-La Mancha	10,8	220.901	16.431	7,4	13,4
9 Cataluña	5,9	451.472	55.144	12,2	8,2
10 C. Valenciana	2,5	125.716	10.740	8,5	11,7
11 Extremadura	3,0	31.853	2.919	9,2	10,9
12 Galicia	2,1	56.747	9.219	16,2	6,2
13 Madrid	11,3	762.459	64.972	8,5	11,7
14 Murcia	1,4	21.064	1.517	7,2	13,9
15 Navarra	5,8	38.076	5.081	13,3	7,5
16 País Vasco	4,0	87.572	13.176	15,0	6,6
17 La Rioja	3,3	10.426	4.008	38,4	2,6
18 Ceuta	1,1	924	114	12,3	8,1
19 Melilla	1,9	1.605	119	7,4	13,5
ESPAÑA	5,0	2.366.499	227.958	9,6	10,4

Tabla 3

Área bajo la curva ROC de la tasa de incidencia a 7 días el 15 de marzo de 2020, punto de corte óptimo de la curva y medidas de precisión asociadas al punto de corte para 2 niveles de alerta

	Nivel 1: evolución intermedia		Nivel 2: peor evolución	
	Valor	IC95%	Valor	IC95%
Área bajo la curva ROC	0,770	0,643-0,898	0,757	0,569-0,945
Punto de corte óptimo	134,5		166,9	
Sensibilidad	59,3%	38,8-77,6	70,0%	34,8-93,3
Especificidad	88,0%	68,8-97,5	81,0%	65,9-91,4
Valor predictivo positivo	84,2%	60,4-96,6	46,7%	21,3-73,4
Valor predictivo negativo	66,7%	48,2-82	91,9%	78,1-98,3

En negrita, puntos de corte.

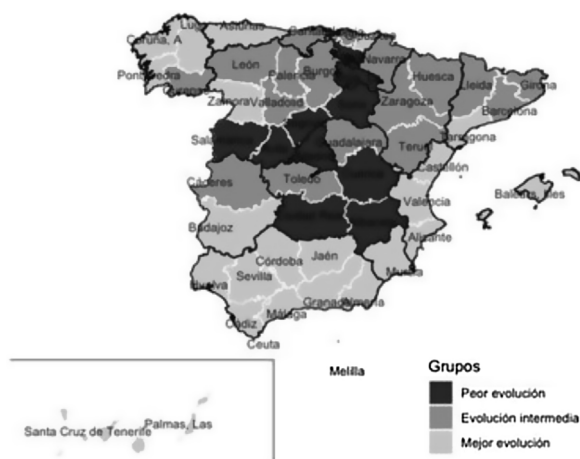


Figura 4. Distribución geográfica de los grupos de diferente evolución de la pandemia.

resultan de tomar como marcador la tasa corregida y como estatus el grupo de peor evolución o evolución intermedia. El área bajo la curva ROC fue 0,757 y 0,770, respectivamente y los puntos de corte en la tasa de incidencia a 7 días fueron 134 casos por 100.000 para el nivel de alerta 1 (medio) y 167 para el nivel 2 (alto); en ambos casos, la especificidad asociada a estos puntos de corte es más alta que la sensibilidad (tabla 3).

Discusión

Los resultados de esta investigación ponen de manifiesto que fijar determinados umbrales en forma de tasas de incidencia a 7 días podría predecir la evolución posterior de la epidemia de SARS-CoV-2 en zonas geográficas concretas. Esos umbrales indican que se puede discriminar adecuadamente con un bajo margen de error entre zonas geográficas de peor evolución, de evolución intermedia y de buena evolución, cuando existe transmisión comunitaria de SARS-CoV-2. La metodología propuesta podría ser útil para otros países, aunque previamente sería necesario hacer estudios similares para comprobar su utilidad y si se mantiene el grado de discriminación. Que sepamos, este es el primer estudio que aplica esta metodología para predecir el comportamiento de la pandemia basado en la incidencia de la enfermedad.

En España y en otros países, ante una situación nueva y desconocida hasta la fecha, se aplicaron medidas de confinamiento extraordinarias. Los resultados de este estudio indican que, con esos umbrales, el confinamiento podría haber llegado tarde en algunas provincias y que en otras quizá habría podido ser evitado o acortado. Si se comparan estos resultados con los observados en el estudio nacional ENE-COVID19⁹, puede observarse que, de las 10 provincias identificadas en nuestro estudio como de peor evolución, 6 coinciden con las que han obtenido más de un 10% de incidencia en dicho estudio. De las 23 provincias identificadas en nuestro estudio como de mejor evolución, 16 presentaron una incidencia menor del 2,5% según el estudio ENE-COVID19. No obstante, debe destacarse que ambos estudios no son comparables. Mientras el presente estudio trata de predecir la buena o mala evolución

y clasificarla en diferentes escenarios, el estudio ENE-COVID19 tiene como objetivo conocer la incidencia de SARS-CoV-2 e indirectamente la prevalencia por provincia con el objetivo último de obtener información sobre la inmunidad de grupo. Así, la propuesta que presentamos mide también la velocidad de infección por SARS-CoV-2, que es uno de los 2 componentes que explican el 99% de la variabilidad y que tiene un impacto directo en la peor evolución. Se puede alcanzar una misma incidencia acumulada de enfermedad al cabo de 2 meses en una misma provincia, pero a través de diferentes mecanismos de transmisión, uno que podría ser lento y sostenido en el tiempo (aplanar la curva) y otro explosivo que satura el sistema sanitario, equivalente al escenario de peor evolución. La combinación de ambos componentes, siendo el primero la tendencia global de la incidencia acumulada por cada provincia y el segundo la velocidad de crecimiento o decrecimiento de la incidencia a lo largo del período explica la práctica totalidad de la evolución de la pandemia en cada provincia analizada, tal y como se puede observar en la [figura 2](#).

Si bien puede parecer que los umbrales propuestos no tienen en cuenta el confinamiento y que algunas provincias alcanzaron el pico de contagios posteriormente al Decreto de Estado de Alarma, para definir las curvas ROC se ha considerado la tasa acumulada a 7 días pero calculada a 15 de marzo y el área bajo la curva es superior a 0,7 en los 2 casos, con lo que los umbrales establecidos de 134 y 167 casos por 100.000 habitantes son capaces de discriminar el nivel de alerta medio y alto, respectivamente.

A la hora de aplicar estos modelos en nuevos contextos como en la situación actual deben hacerse varias consideraciones. En primer lugar, el modelo es fundamentalmente aplicable en situaciones en las que existe transmisión comunitaria, de origen desconocido y no trazable. En situaciones en las que el número de casos, aunque sea elevado, es de origen conocido y pueden trazarse sus contactos, la velocidad de transmisión puede ser muy alta, pero durante un tiempo muy limitado, por lo que no podría hablarse de una mala evolución sino de un aumento puntual y localizado del número de casos. Finalmente, se ha decidido eliminar de los modelos los casos confirmados a través de test rápidos debido a que su uso heterogéneo en las distintas provincias españolas podría afectar a los resultados, al igual que cierta variabilidad en su fiabilidad diagnóstica. Debemos destacar además que su sensibilidad y especificidad son menores que la prueba de PCR y que en cualquier caso solo supusieron un 7,7% del total. Por provincias, la variación fue importante, entre el 0% y el 45%, con una mediana del 11,2%.

No hemos podido considerar en estos modelos el número de ingresos hospitalarios por provincia, ingresos en UCI o fallecimientos, debido a que no estaban disponibles. De todos modos, haber incluido información de estas características hubiese supuesto introducir como elemento distorsionador la diferente preparación de los sistemas de atención sanitaria ante la pandemia.

Los umbrales obtenidos en nuestro estudio son comparables a los propuestos por otros organismos. Así, el propuesto por el Harvard Global Health Institute específica, para el nivel máximo de alerta, 25 casos diarios por 100.000 habitantes¹⁵. Nuestro modelo indica como este nivel de alerta 167 casos a 7 días, lo que significa 24 (167/7) casos diarios, prácticamente igual al propuesto por esta institución. Sin embargo, los 2 niveles intermedios calculados no son similares. Hay que destacar que nuestro modelo ha sido propuesto de manera ciega y totalmente independiente a lo propuesto por esta institución académica norteamericana. Existen múltiples estudios de modelización de COVID-19 pero se han centrado en aspectos diversos, como predicción de carga hospitalaria, predicción de mortalidad, etc.¹⁶⁻¹⁸, pero pocos han sido realizados para establecer la utilidad de umbrales epidemiológicos que puedan predecir la evolución de la pandemia. Otro documento de revisión canadiense indica que el umbral que puede predecir capacidad hospitalaria es el de 30-50 casos por 100.000 habitantes¹⁹ (bastante cercano al

corte que proponemos de 24 casos por 100.000), pero la OMS no establece umbrales de acción basados en tasas de incidencia sino en la existencia de transmisión comunitaria o clústers de casos²⁰.

Por otra parte, la principal fortaleza del método estadístico empleado está en que permite hacer la clasificación de las provincias teniendo en cuenta su evolución a lo largo del período epidémico, utilizando para ello toda la información de las curvas de tasas de incidencia por 100.000 habitantes. Otra ventaja es que creemos que este tipo de modelos podrían ser aplicables a esta o futuras pandemias, modelizando los 2 componentes analizados en función de las características epidémicas de ese potencial nuevo agente biológico (tiempo de inducción, infectividad).

Finalmente, en cuanto a las limitaciones del estudio, tenemos que decir que la tasa de detección utilizada para hacer la corrección de las tasas a 7 días calculada el 15 de marzo está estimada con los datos de todo el período hasta el 11 de mayo y se ha asumido que es válida a día 15 de marzo. Además, debido a la falta de datos a nivel provincial, se aplicó a todas las provincias de una CA la tasa de detección de la CA, pero podría haber diferencias interprovinciales. También, las tasas de incidencia empleadas para la clasificación de las provincias son por inicio de síntomas, debido a la no disponibilidad de datos por fecha de diagnóstico, que serían los deseables para este estudio.

Para concluir, los umbrales propuestos para establecer una evolución intermedia o una peor evolución en forma de tasa de incidencia a 7 días podrían ser de utilidad para las administraciones sanitarias a la hora de decidir medidas que puedan afectar a la movilidad de la población, siempre y cuando una zona geográfica se encuentre en una situación de transmisión comunitaria de SARS-CoV-2. Esta información es sencilla de obtener a través de los sistemas de información de Salud Pública de las distintas CCAA y puede ser intercomparable entre áreas sanitarias o CCAA.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

- World Health Organization (WHO). Novel coronavirus (2019-nCoV) situation report – 38, 27 February 2020. WHO Bulletin; 2020 [consultado 25 Jul 2020]. Disponible en: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19. BOE núm. 67, de 14 de Mar de 2020; p. 25390-400 [consultado 25 Jul 2020]. Disponible en: <https://www.boe.es/eli/es/rd/2020/03/14/463>.
- Instituto de Salud Carlos III. Situación y evolución de la pandemia de COVID-19 en España: Panel COVID-19 [consultado 1 Jul 2020]. Disponible en: <https://cnecovid.isciii.es>.
- Boscá JE, Doménech R, Ferri J. COVID-19: hospitalizaciones y presión sobre el sistema sanitario español. BBVA Research [consultado 25 Jul 2020]. Disponible en: <https://www.bbva.com/publicaciones/covid-19-hospitalizaciones-y-presion-sobre-el-sistema-sanitario-espanol/>.
- Instituto de Salud Carlos III. Sistema de Monitorización de la Mortalidad diaria (MoMo) [consultado 25 Jul 2020]. Disponible en: <https://momo.isciii.es/public/momo/dashboard/momo.dashboard.html>.
- Banco de España. Escenarios macroeconómicos de referencia para la economía española tras el COVID-19. Boletín Económico 2/2020 [consultado 25 Jul 2020]. Disponible en: <https://www.bde.es/f/webbde/GAP/Secciones/SalaPrensa/COVID-19/be2002-art1.pdf>.
- Instituto Nacional de Estadística. Cifras de población. Datos provisionales a 1 de Ene de 2020 [consultado 1 Jul 2020]. Disponible en: <https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica.C&cid=1254736176951&menu=ultiDatos&idp=1254735572981>.
- Escovid19data: Análisis de COVID-19 por provincias en España [consultado 1 Jul 2020]. Disponible en: <https://lab.montera34.com/covid19/provincias.html>.
- Instituto de Salud Carlos III. (2020). Estudio nacional de sero-epidemiología de la infección por SARS-CoV-2 en España (ENECOVID) [consultado 1 Nov 2020]. Disponible en: <https://portalcne.isciii.es/enecovid19/>.
- Silverman BW, Ramsay JO. Functional data analysis. 2nd ed. Nueva York: Springer; 2005.

11. Febrero-Bande M, Oviedo de la Fuente M. Statistical Computing in Functional Data Analysis: The R Package fda.usc. *J Stat Soft.* 2012;51:1–28. Disponible en: <http://www.jstatsoft.org/v51/i04/> [consultado 2 Jul 2020].
12. Pérez-Montilla A. Análisis de datos funcionales: una aplicación en componentes principales [consultado 2 Jul 2020]. Disponible en: <https://rodin.uca.es/xmlui/handle/10498/18510>.
13. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: Cluster analysis basics and extensions. R package version 2.1.0; 2019.
14. Instituto de Salud Carlos III. Informes COVID-19 [consultado 1 Nov 2020]. Disponible en: <https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx>.
15. Harvard Global Health Institute. Key Metrics for COVID Suppression. Framework for policy makers and the public. July, 1 2020 [consultado 7 Nov 2020]. Disponible en: <https://globalepidemics.org/key-metrics-for-covid-suppression/>.
16. Barbarossa MV, Fuhrmann J, Meinke JH, Krieg S, Varma HV, Castelletti N, et al. Modeling the spread of COVID-19 in Germany: Early assessment and possible scenarios. *PLoS One.* 2020;15:e0238559, <http://dx.doi.org/10.1371/journal.pone.0238559>.
17. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med.* 2020;26:855–60, <http://dx.doi.org/10.1038/s41591-020-0883-7>.
18. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science.* 2020;369:eabb9789, <http://dx.doi.org/10.1126/science.abb9789>.
19. COVID-19 Scientific Advisory Group Rapid Evidence Report [consultado 11 Nov 2020]. Disponible en: <https://www.albertahealthservices.ca/assets/info/ppih/if-ppih-covid-19-sag-models-scenarios-and-thresholds-rapid-review.pdf>.
20. Critical preparedness, readiness and response actions for COVID-19. Interim guidance. March, 22 2020 [consultado 11 Nov 2020]. Disponible en: <https://apps.who.int/iris/bitstream/handle/10665/331511/Critical%20preparedness%20readiness%20and%20response%20actions%20COVID-10%202020-03-22.FINAL-eng.pdf?sequence=1&isAllowed=y>.