



**ORIGINAL BREVE**

## Estudio de la incidencia de COVID-19 en España y su relación geográfica provincial



J.J. Alfaro-Martínez<sup>a,\*</sup>, J. Solís García del Pozo<sup>b</sup>, R.P. Quílez Toboso<sup>a</sup>, L. García Blasco<sup>a</sup>  
y C. Rosa Felipe<sup>c</sup>

<sup>a</sup> Servicio de Endocrinología y Nutrición, Complejo Hospitalario Universitario de Albacete, Albacete, España

<sup>b</sup> Unidad de Enfermedades Infecciosas, Complejo Hospitalario Universitario de Albacete, Albacete, España

<sup>c</sup> Servicio de Medicina Interna, Complejo Hospitalario Universitario de Albacete, Albacete, España

Recibido el 21 de septiembre de 2022; aceptado el 1 de febrero de 2023

Disponible en Internet el 14 de febrero de 2023

### PALABRAS CLAVE

COVID-19;  
Incidencia;  
Epidemiología;  
España;  
Conglomerados;  
Provincias;  
Comunidad autónoma

### Resumen

**Introducción:** El objetivo de este trabajo fue conocer la incidencia de COVID-19 en España en relación con la geografía en provincias durante el conjunto de la pandemia.

**Método:** Análisis de conglomerados de la incidencia de COVID-19 en las provincias y ciudades españolas en las seis primeras olas de la pandemia.

**Resultados:** Todas las provincias de Canarias, Cataluña y Andalucía forman conglomerados sin presencia de ninguna otra provincia. Además, en la Comunidad Valenciana, Galicia, País Vasco y Aragón dos de las tres provincias (tres de cuatro en el caso de Galicia) estaban en el mismo conglomerado.

**Discusión:** La incidencia de COVID-19 en las seis primeras olas forma conglomerados que reproducen la división de España en comunidades autónomas. Aunque esto podría explicarse por una mayor movilidad dentro de una comunidad, no puede descartarse que esta distribución se deba a diferencias a la hora de cribar, diagnosticar, registrar o comunicar casos de COVID-19.

© 2023 FECA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

### KEYWORDS

COVID-19;  
Incidence;  
Epidemiology;  
Spain;  
Clusters;  
Provinces;  
Autonomous  
communities

**Study of incidence of COVID-19 in Spain and its relationship to geographical province distribution**

### Abstract

**Background:** The aim of this study was to determine incidence of COVID-19 in relationship to geographical distribution among Spain.

**Method:** Cluster analysis taking into consideration the incidence of COVID-19 in the provinces and autonomous cities of Spain in each of the first six waves of the pandemic.

\* Autor para correspondencia.

Correo electrónico: [josej@alfaro.nom.es](mailto:josej@alfaro.nom.es) (J.J. Alfaro-Martínez).

**Results:** All the provinces of the Canary Islands, Catalonia and Andalusia form independent clusters. In Comunidad Valenciana, Galicia, País Vasco and Aragón two out of three provinces (three out of four in Galicia) were in the same cluster, with no other provinces.

**Discussion:** The incidence of COVID-19 in Spain in the first six waves forms clusters that reproduce the territorial division of Spain into autonomous communities. Although this could be explained by greater mobility within a community, it cannot be ruled out that this distribution is due to differences in screening, diagnosis, registration or reporting of COVID-19 cases.

© 2023 FECA. Published by Elsevier España, S.L.U. All rights reserved.

## Introducción

Desde que se diagnosticó el primer caso de COVID-19 en España hasta que dejó de facilitarse información sobre contagios en toda la población se han sucedido en nuestro país seis olas de contagios. Los datos sobre contagios, hospitalización y defunciones, hasta un nivel provincial, han estado libremente disponibles y han servido para realizar análisis referidos a determinadas olas y territorios, por ejemplo el de Santamaría y Hortal<sup>1</sup>.

A lo largo del tiempo la evolución de la pandemia en España no ha sido uniforme en los distintos territorios. La primera ola golpeó de forma especialmente severa a provincias como Madrid o Albacete, entre otras, mientras que otras la sufrieron con menor virulencia, aunque se vieron más afectadas por sucesivas olas<sup>2,3</sup>.

Sin embargo, no conocemos estudios sobre la relación de unos territorios con otros en cuanto a la incidencia de COVID-19 en las distintas olas.

El objetivo de este trabajo fue analizar, a nivel provincial, la incidencia de COVID-19 en las seis primeras olas de la pandemia en España para conocer qué provincias tuvieron un comportamiento similar a otras durante el conjunto de la pandemia.

## Material y métodos

Se realizó un análisis de conglomerados de la incidencia de COVID-19 en las provincias y ciudades autónomas de España en cada una de las seis primeras olas de la pandemia. Para ello se calculó la incidencia de COVID-19 en cada provincia y ola según los datos de contagios proporcionados por el Instituto de Salud Carlos III<sup>4</sup> y los datos de población por provincias en el año 2021<sup>5</sup>.

De forma arbitraria, se emplearon los siguientes límites de tiempo entre una ola y otra: primera ola hasta el 31 de julio de 2020, segunda ola hasta el 30 de noviembre de 2020, tercera ola hasta 31 de marzo de 2021, cuarta ola hasta 14 de junio de 2021, quinta ola hasta el 30 de noviembre de 2021 y sexta ola hasta el 31 de marzo de 2022. Estos límites se basaron en la gráfica de incidencia de toda España obtenida de los datos del Instituto Carlos III, marcando la separación entre una ola y otra en el punto en que se observaba un cambio de tendencia.

Con la incidencia de cada ola y provincia/ciudad autónoma se calculó la matriz de distancias euclídeas entre cada par de provincias/ciudades autónomas mediante la función `dist()` del lenguaje estadístico R<sup>6</sup> versión 4.1.2.

La matriz de distancias contiene información de lo diferente que fue cada par de provincias/ciudades autónomas en incidencia en cada una de las seis olas. Si solo se estudiara una ola la distancia entre dos entidades, sería la diferencia de incidencias. Si se estudiaran dos olas podríamos representar, para cada par de entidades, en un sistema de coordenadas la incidencia en la primera ola en el eje de las x y la de la segunda en el eje de las y, y la distancia euclídea sería la distancia entre los dos puntos «en diagonal». Si estudiáramos tres olas podríamos representar las incidencias de cada ola en los tres ejes x, y, z de un sistema de coordenadas en tres dimensiones, y medir la distancia entre los puntos de cada par de entidades. Para las seis olas sería lo mismo pero en un espacio de 6 dimensiones.

Partiendo de la matriz de distancias se pueden formar conglomerados (es decir, grupos de entidades que están más cerca entre sí que de otras entidades que a su vez se pueden agrupar) mediante diferentes técnicas estadísticas, cuya explicación se sale del objeto de este trabajo<sup>7</sup>.

En nuestro trabajo se formaron conglomerados, basándose en la matriz de distancias euclídeas arriba referida, mediante la técnica de Ward, mediante la función `hclust()` del lenguaje estadístico R versión 4.1.2, indicando como parámetro `method = «ward.D»`.

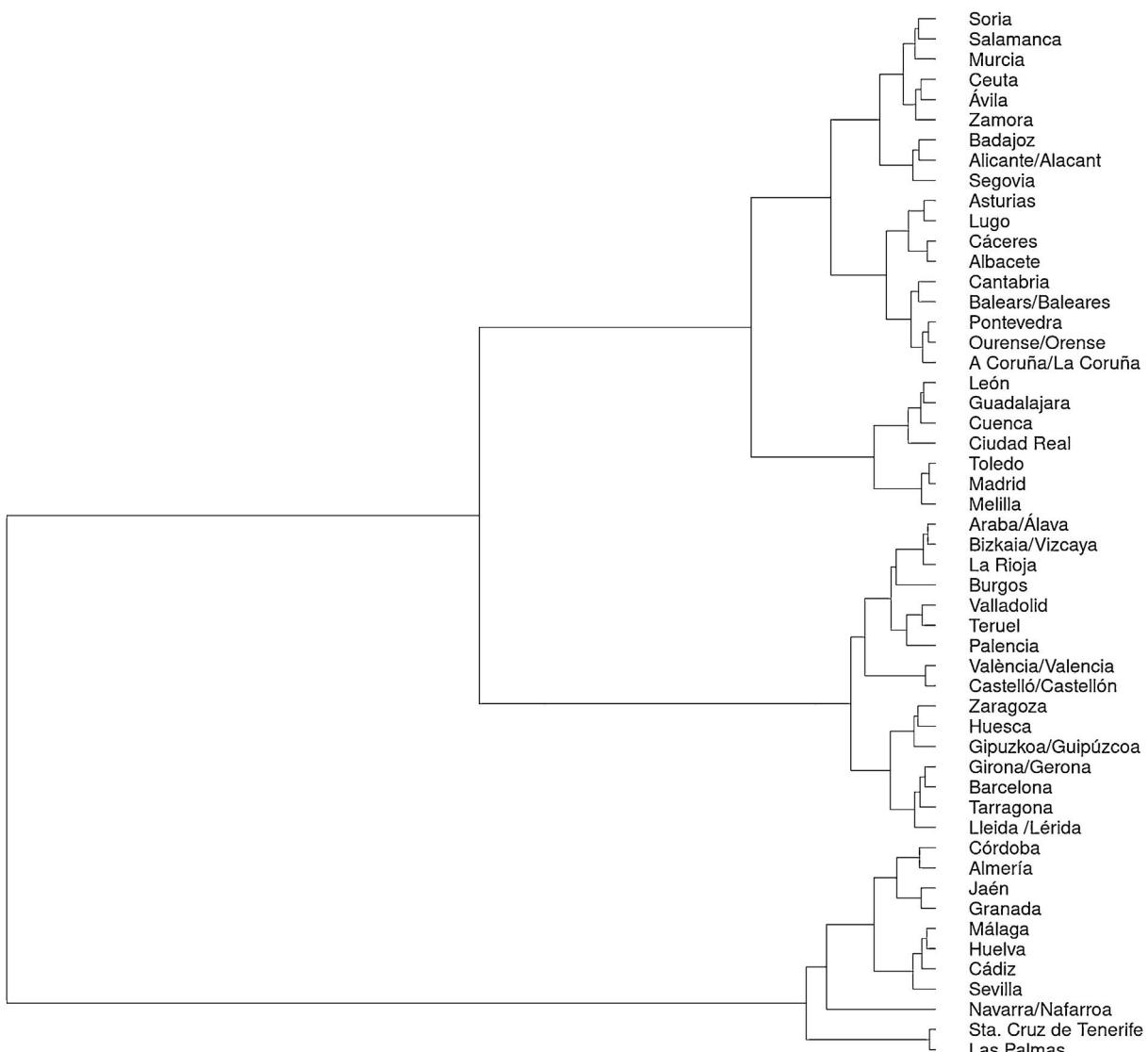
Finalmente se representaron los conglomerados obtenidos mediante las funciones `as.dendrogram()` y `plot()` del mencionado lenguaje estadístico.

Adicionalmente se formaron conglomerados con otras técnicas (single, complete, average, centroid, Mcquitty y median) utilizando igualmente la función `hclust()` con el parámetro `method` correspondiente.

Finalmente, siguiendo la misma técnica, se llevó a cabo un análisis de conglomerados de la tasa de hospitalizaciones y mortalidad por COVID-19 en las seis olas en las diferentes provincias y ciudades autónomas de España.

## Resultados

El resultado del análisis de conglomerados se muestra en la figura 1.



**Figura 1** Análisis de conglomerados de la incidencia de COVID-19 en las seis primeras olas en las provincias y ciudades autónomas de España. Puede observarse cómo muchas provincias que forman parte de comunidades autónomas multiprovinciales se agrupan junto con otras provincias de su misma comunidad en conglomerados en los que no hay provincias de otras comunidades. Por ejemplo, en la parte inferior del gráfico, y de abajo arriba, vemos las dos provincias de Canarias en un conglomerado, tras Navarra, las ocho provincias andaluzas en otro, y las cuatro provincias catalanas en un tercero.

Analizando las diez comunidades autónomas multiprovinciales, se observó que todas las provincias de Canarias, Cataluña y Andalucía forman conglomerados sin presencia de ninguna otra provincia ajena a la Comunidad. Además, en las comunidades autónomas de Comunidad Valenciana, Galicia, País Vasco y Aragón, dos de las tres provincias (tres de cuatro en el caso de Galicia) estaban en el mismo conglomerado, sin otra provincia ajena a la comunidad autónoma.

La [tabla 1](#) muestra el resultado de la formación de conglomerados por otros métodos. En general, independientemente del método utilizado, las provincias de las comunidades autónomas multiprovinciales arriba mencionadas tendían a formar conglomerados sin presencia de otras provincias.

Solo en el caso de Extremadura y de Castilla y León las provincias de una misma comunidad autónoma no tienden

a agruparse. Las provincias extremeñas nunca están en el mismo conglomerado sea cual sea el método utilizado, y solo dos de las de Castilla y León coinciden. Las provincias de Castilla-La Mancha, salvo Albacete, tienden a estar en el mismo conglomerado, junto con Madrid y León.

Al contrario que con la incidencia de casos, el análisis de conglomerados de la tasa por provincias y olas de ingresos hospitalarios y mortalidad por COVID-19 mostró una tendencia mucho menor a formar conglomerados de las provincias de comunidades autónomas multiprovinciales.

## Discusión

Nuestro trabajo muestra que la incidencia de COVID-19 en las distintas provincias de España en las seis primeras olas

**Tabla 1** Número de provincias pertenecientes a cada comunidad autónoma multiprovincial que son agrupadas en el mismo conglomerado, al analizar incidencias provinciales y por olas, sin que en el mismo haya provincias de otra comunidad autónoma, en función de la técnica estadística para la formación de conglomerados, en relación con el total de provincias de la comunidad autónoma

	Single	Complete	Average	Centroid	Mcquitty	Median	Ward
Andalucía	8/8	8/8	8/8	8/8	8/8	8/8	8/8
Cataluña	2/4	4/4	4/4	3/4	4/4	3/4	4/4
Castilla-La Mancha	0/5	0/5	0/5	0/5	0/5	0/5	0/5
Castilla y León	0/9	2/9	2/9	2/9	2/9	0/9	2/9
Canarias	2/2	2/2	2/2	2/2	2/2	2/2	2/2
Extremadura	0/2	0/2	0/2	0/2	0/2	0/2	0/2
Valencia	2/3	2/3	2/3	2/3	2/3	2/3	2/3
Galicia	3/4	3/4	3/4	3/4	3/4	3/4	3/4
País Vasco	2/3	2/3	2/3	2/3	2/3	2/3	2/3
Aragón	0/3	2/3	2/3	0/3	2/3	0/3-	2/3

forma conglomerados, los cuales siguen de forma fiel la división territorial de España en comunidades autónomas. No así la tasa de hospitalizaciones o la de defunciones por COVID-19, que han sido analizados en otros trabajos a nivel autonómico<sup>8</sup>, y que en nuestro caso, analizándolo a nivel provincial, no forma conglomerados que reproduzcan la división autonómica de España.

Aunque no hay una definición oficial sobre el número de olas de COVID-19 que ha habido en España, y la separación entre ellas, los límites que hemos empleado son similares, en algún caso idénticos, a los de otros trabajos<sup>9</sup>, y como la separación entre olas corresponde a períodos de incidencia baja, unos días de diferencia en el límite entre olas no afecta significativamente a la incidencia del conjunto de la ola.

La distribución de las provincias en conglomerados que mimetizan comunidades autónomas podría deberse a que entre provincias limítrofes, y más si son de la misma comunidad, puede haber más movilidad, y también a las restricciones a los desplazamientos entre comunidades autónomas que se establecieron durante el estado de alarma<sup>3,10</sup>. Sin embargo, provincias limítrofes de distintas comunidades aparecen en conglomerados que están separados (salvo Madrid y las provincias de Castilla-La Mancha), y las restricciones no han afectado a las seis olas. El análisis de conglomerados agrupa provincias con una incidencia similar en todas y cada una de las seis olas, por lo que restricciones (o también salidas vacacionales) de una sola ola no tendrían mucha repercusión en la formación de conglomerados. Las provincias, como Albacete o León, que no están en el mismo conglomerado que el resto de las provincias de su comunidad pueden haber tenido distinto comportamiento que estas en una o más olas por motivos que desconocemos. Diversos trabajos han relacionado factores geográficos o climatológicos con la heterogeneidad de la incidencia de COVID-19 en España<sup>11-13</sup>, pero geografía y clima no siguen de forma tan fiel los límites políticos de las provincias y comunidades autónomas de España.

Por otra parte, cuando complementariamente se analizan por conglomerados las hospitalizaciones o la mortalidad por COVID-19, estos ya no reflejan de una forma tan fiel la

división administrativa de España. El hecho de que en el análisis de las hospitalizaciones y la mortalidad los resultados sean distintos puede deberse a que la estructura poblacional de las provincias dentro de una misma comunidad es distinta, por lo que con una misma incidencia las consecuencias clínicas no son iguales. O también puede deberse a que mortalidad u hospitalizaciones son variables más «duras», menos dependientes de los criterios con los que se efectúan test diagnósticos ambulatorios.

En definitiva, partiendo de los datos de incidencia provincial de COVID-19 no podemos conocer la causa de este agrupamiento en conglomerados, tan solo aventurar suposiciones. Si bien una mayor movilidad dentro de una comunidad podría explicar los conglomerados, no puede descartarse que esta distribución se deba a diferencias entre unas comunidades autónomas y otras a la hora de cribar, diagnosticar, registrar o comunicar casos de COVID-19, lo que habría de tenerse en cuenta a la hora de interpretar los datos de incidencia en España.

El análisis de conglomerados empleado en este trabajo es una técnica estadística exploratoria, no contrasta hipótesis, no permite asignar un umbral de significación estadística a partir del cual dos elementos están en distinto conglomerado. Sin embargo, el hecho de que en nuestro trabajo los resultados obtenidos respecto a las comunidades autónomas multiprovinciales cambien poco independientemente del método utilizado indica que los resultados obtenidos son robustos<sup>7</sup>.

En conclusión, el análisis de conglomerados indicaría que la evolución de la incidencia de COVID-19 en las seis primeras olas fue similar en las provincias pertenecientes a una comunidad autónoma, excepto en el caso de las provincias extremeñas y castellano-leonesas, aunque este tipo de estudios no permite conocer la causa de ello.

## Financiación

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades sin ánimo de lucro.

## Conflicto de intereses

Ninguno de los autores presenta conflicto de intereses en relación con el presente trabajo.

## Bibliografía

1. Santamaría L, Hortal J. COVID-19 effective reproduction number dropped during Spain's nationwide dropdown, then spiked at lower-incidence regions. *Sci Total Environ.* 2021;751:142257, <http://dx.doi.org/10.1016/j.scitotenv.2020.142257>.
2. Briz-Redón A, Serrano-Aroca A. On the association between COVID-19 vaccination levels and incidence and lethality rates at a regional scale in Spain. *Stoch Environ Res Risk Assess.* 2022;36:2941–8, <http://dx.doi.org/10.1007/s00477-021-02166-y>.
3. Orea L, Álvarez IC. How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces. *Health Econ.* 2022;31:154–73, <http://dx.doi.org/10.1002/hec.4437>.
4. CNE-ISCIII. Casos por técnica y provincia [consultado 1 Abr 2022]. Disponible en: [https://cnecovid.isciii.es/covid19/resources/casos\\_tecnica\\_provincia.csv](https://cnecovid.isciii.es/covid19/resources/casos_tecnica_provincia.csv).
5. Instituto Nacional de Estadística. Población por provincias y sexos [consultado 1 Abr 2022]. Disponible en: <https://www.ine.es/jaxiT3/Tabla.htm?t=2852>.
6. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Disponible en: <https://www.R-project.org/>.
7. García Pérez A. Análisis de conglomerados. En: *Métodos avanzados de estadística aplicada. Técnicas avanzadas.* 1.<sup>a</sup> edición UNED; 2005.
8. Ruiz-García A, Vitelli-Storelli F, Serrano-Cumplido A, Segura-Fragoso A, Calderón-Montero A, Mico-Pérez RM, et al. Tasas de letalidad por SARS-CoV-2 según comunidades autónomas durante la segunda onda epidémica en España. *Semergen.* 2022;48:252–62, <http://dx.doi.org/10.1016/j.semerg.2022.02.004>.
9. Pilo de la Fuente B, González Martín-Moro J, Martín Ávila G, Méndez Burgos A, Ramos Barrau L, Thuissard I, et al. Risk and prognostic factors for SARS-CoV-2 infection in Spanish population with multiple sclerosis during the first five waves. *Front Neurol.* 2022;13:1001429, <http://dx.doi.org/10.3389/fneur.2022.1001429>.
10. Mazzoli M, Mateo D, Hernando A, Meloni S, Ramasco JJ. Effects of mobility and multi-seeding on the propagation of the COVID-19 in Spain. *medRxiv.* 2020, <http://dx.doi.org/10.1101/2020.05.09.20096339>.
11. Valero C, Barba R, Marcos DP, Puente N, Riancho JA, Santurtún A. Influence of weather factors on the incidence of COVID-19 in Spain. *Med Clin (Engl Ed).* 2022;159:255–61, <http://dx.doi.org/10.1016/j.medcle.2021.10.018>.
12. Martorell-Marugán J, Villatoro-García JA, García-Moreno A, López-Domínguez R, Requena F, Merelo JJ, et al. DataC: A visual analytics platform to explore climate and air quality indicators associated with the COVID-19 pandemic in Spain. *Sci Total Environ.* 2021;750:141424, <http://dx.doi.org/10.1016/j.scitotenv.2020.141424>.
13. Shahzad K, Shahzad U, Iqbal N, Shahzad F, Fareed Z. Effects of climatological parameters on the outbreak spread of COVID-19 in highly affected regions of Spain. *Environ Sci Pollut Res.* 2020;27:39657–66, <http://dx.doi.org/10.1007/s11356-020-10551-3>.