Informe breve 17-06-2020

# Análisis de los retrasos en la actualización de las series históricas de casos en España

Martí Català, Sergio Alonso, Enric Álvarez, Daniel López, Miquel Marchena, David Conesa, Pere-Joan Cardona, Clara Prats

*Comparative Medicine and Bioimage Centre of Catalonia; Institute for Health Science Research Germans Trias i Pujol*
*Computational Biology and Complex Systems; Universitat Politècnica de Catalunya - BarcelonaTech*

**Con la colaboración de:** Guillem Álvarez, Oriol Bertomeu, Laura Dot, Lavínia Hriscu, Helena Kirchner, Miquel Marchena, Daniel Molinuevo, Pablo Palacios, Sergi Pradas, David Rovira, Xavier Simó, Tomás Urdiales

Contacto: clara.prats@upc.edu

*With the financial support of:* MINISTERIO DE CIENCIA E INNOVACIÓN "la Caixa"

## Introducción

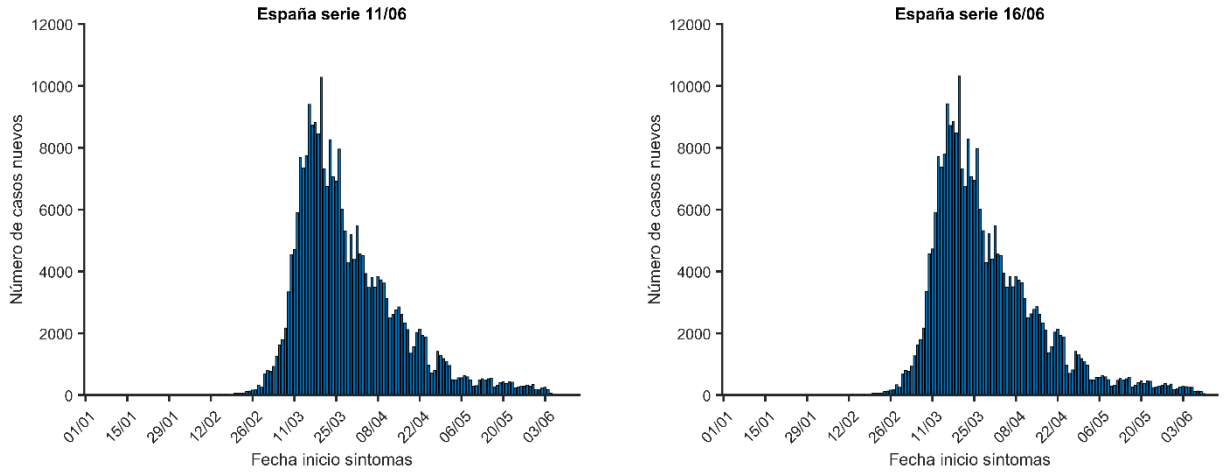
Este informe breve tiene como objetivo evaluar los retrasos presentes en las actualizaciones de las series históricas publicadas por el ISCiii. Desde el 10 de junio de 2020, dicho organismo publica periódicamente una revisión actualizada de las series históricas de casos de Covid-19 para las provincias y comunidades autónomas españolas. En estas series, la fecha de referencia para los casos es la de aparición de los síntomas. En los casos asintomáticos, se asignan a 6 días antes de la fecha de diagnóstico. Es de esperar que este criterio conduzca a un **retraso aparente** en las notificaciones con respecto a los datos publicados con la fecha de diagnóstico o con la propia fecha de notificación. El motivo es que **los casos que han empezado a mostrar síntomas en los últimos días probablemente aún no hayan sido diagnosticados, registrados o validados debido a los retrasos propios del proceso**. Dichos casos serán asignados al día correspondiente en actualizaciones posteriores.

Este retraso en la actualización de los datos correspondientes a los últimos días, probablemente inevitable en muchos de los casos, tiene que ser tenida en cuenta a la hora de analizar dichas series. Por ejemplo, la estimación de la IA7 o de la Rt sólo podrá hacerse hasta el día en que los datos estén más o menos consolidados, esperándose pocas variaciones en actualizaciones sucesivas. En cambio, los datos de los últimos días estarán infraestimando el número total de casos.

El objetivo de este documento es hacer una primera estimación de dicho retraso, **comparando los datos publicados el 11 de junio con los datos publicados el 16 de junio**.

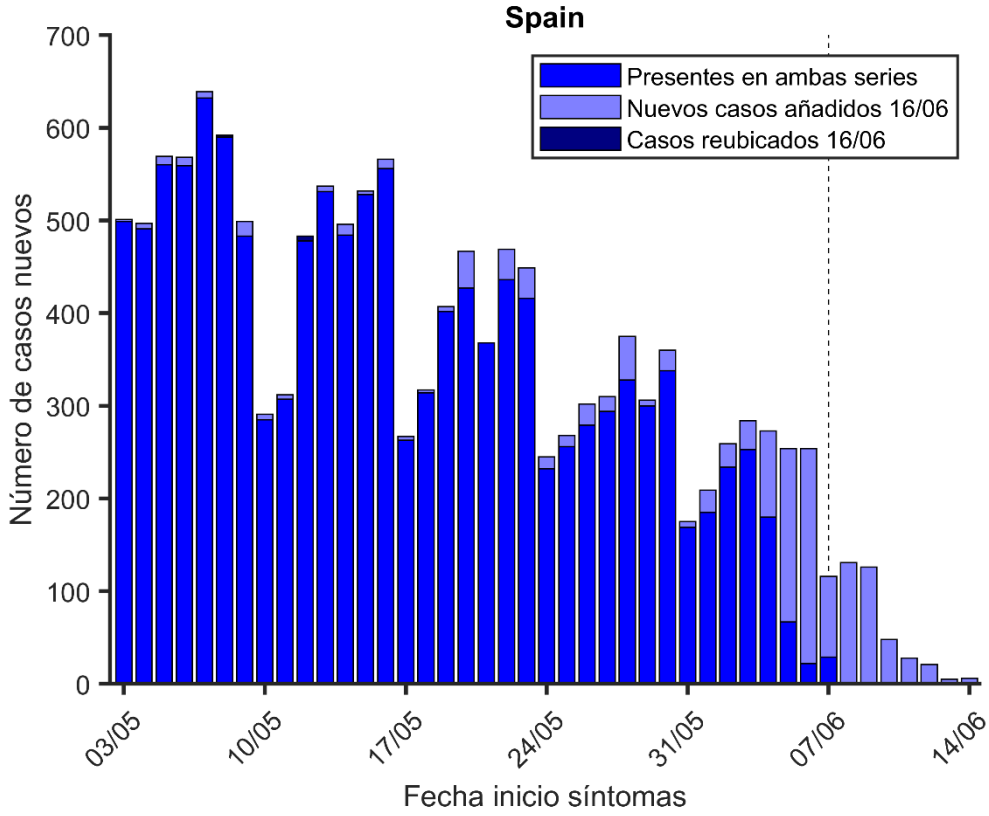
## Análisis global a nivel de país

En la siguiente figura se muestra cómo se distribuyen en el tiempo los casos nuevos, por fecha de diagnóstico, a nivel de España[1]. A la izquierda, serie publicada el 11 de junio de 2020, con datos que llegan hasta el 7 de junio de 2020 (***primera serie***, de aquí en adelante). A la izquierda, serie publicada el 16 de junio de 2020, con datos que llegan hasta el 14 de junio de 2020 (***segunda serie***, de aquí en adelante).




**La segunda serie**, además de aportar una semana más de datos, **modifica la primera en algunos puntos**. En la siguiente gráfica, mostramos los valores presentes en ambas series, así como los nuevos casos añadidos por la segunda serie (azul claro) y los casos que han sido reubicados por la segunda serie (azul oscuro). Se muestran sólo las últimas seis semanas, para poder percibir correctamente los cambios. La línea punteada muestra el día en que acaba la primera serie (7 de junio de 2020).

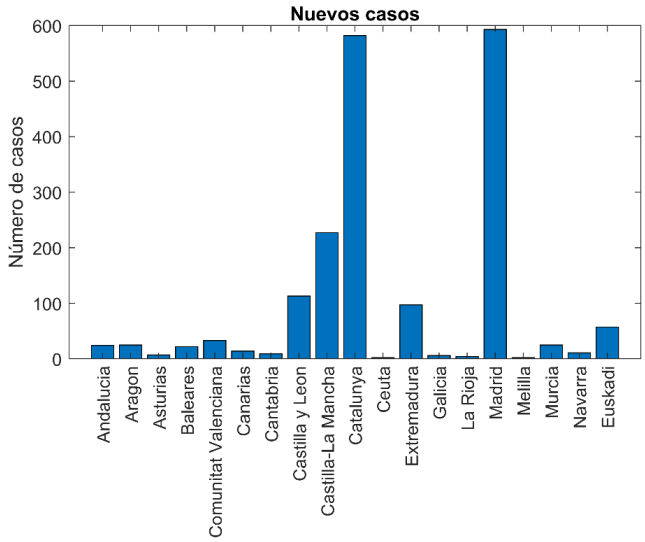
[1] https://cnecovid.isciii.es/covid19/



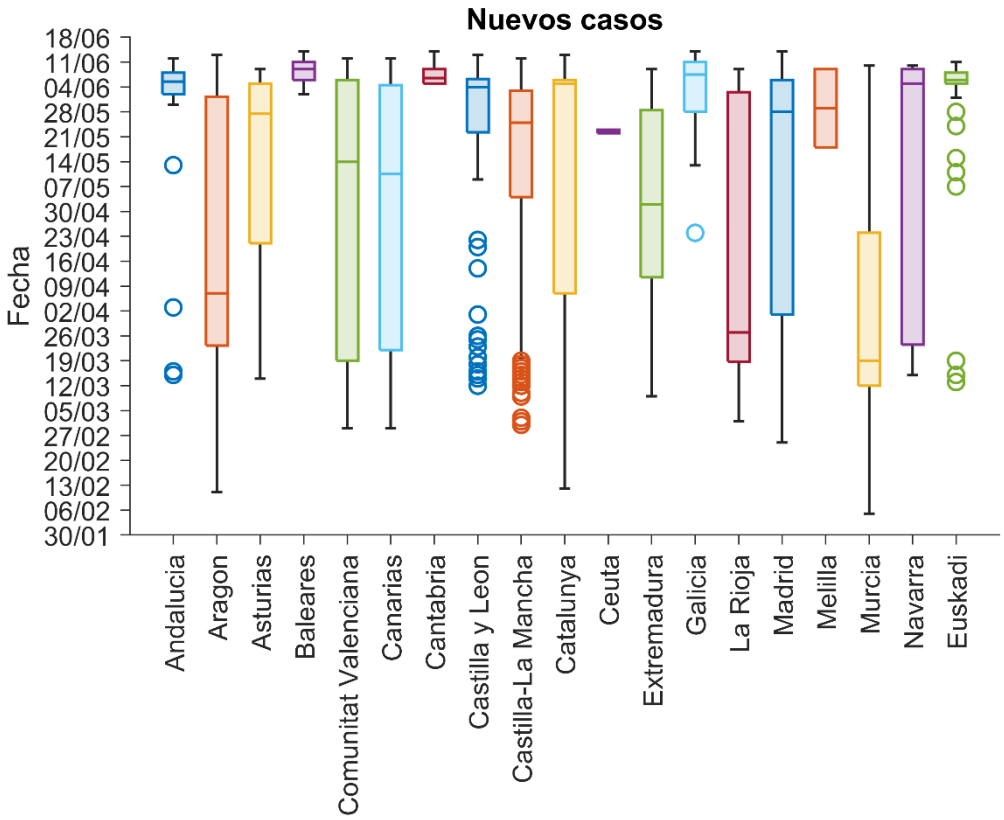

Vemos que la segunda serie modifica algunos valores de días intermedios, en menor o mayor medida según el día. En particular, **observamos que los últimos 4 días de la primera serie son modificados de forma muy significativa por la segunda serie, y los 3 anteriores son modificados en menor medida**. Por último, vemos que los casos nuevos recientes aportados por la segunda serie (última semana, después de la línea punteada) son pocos. De hecho, esperamos que estos valores se vean incrementados con actualizaciones sucesivas. Así, indicadores habitualmente utilizados en el análisis de la situación como el número reproductivo o la incidencia acumulada los últimos 7 días tienen que evaluarse con precaución, ya que *de facto* estarían subestimando el nivel de riesgo epidemiológico real.

## Análisis a nivel de comunidad autónoma

Los cambios detectados a nivel de país varían según la comunidad autónoma. La siguiente figura muestra el total de nuevos casos que añade la segunda serie respecto la primera, para cada comunidad autónoma.



Es interesante ver dónde se sitúan estos casos que añade la segunda serie, en el tiempo. Si no hubiera retrasos, se esperaría que todos los casos que aporta de más la segunda serie correspondieran a la última semana (8 de junio a 14 de junio). No obstante, vemos que se añaden casos en días anteriores, es decir**, se están modificando datos reportados por la primera serie**.




De hecho, **en esta gráfica se combinan dos efectos**: el **retraso en el diagnóstico, registro y validación de los casos de los últimos días** con **procesos de revisión más profunda que están afectando a la consolidación de las series históricas**. Los efectos de esta revisión se perciben de forma especial en comunidades donde se están modificando mayoritariamente datos de marzo, abril y principios de mayo.

En el apéndice A se muestra la comparación temporal de las dos series de datos comunidad a comunidad para las últimas 6 semanas. En general, se observa que las comunidades autónomas presentan cambios significativos en los últimos 3 o 4 días, en la mayoría de ellas. Catalunya y Castilla y León presentarían entre 4 y 5 días de retraso, y Andalucía se situaría en 6. Por último, Castilla la Mancha presenta modificaciones en la segunda serie que van más atrás en el tiempo, aunque podría ser un tema circunstancial del proceso de validación al que se están sometiendo los datos.

## Evaluación del error acumulado

Para evaluar el error cometido, se ha utilizado la metodología siguiente. Se toma como inicio el 7 de junio, y se estudia la serie en sentido inverso, desde el final hasta el principio. Para cada día se evalúa el porcentaje de error acumulado como:

$$Error\ acumulado(t) = \frac{\sum_{t_f}^{t_f - t} Diferencia\ de\ casos\ diarios\ entre\ ambas\ series}{\sum_{t_f}^{t_f - t} Casos\ nuevos\ reportados\ por\ la\ segunda\ serie}$$

Este error se evalúa para toda la serie histórica empezando a $t_f = 7\ de\ junio$. Para cada región (comunidad autónoma y país), se evalúa si el error acumulado está por debajo del 10 % en toda la serie. Si no es así, se va un día atrás y se hace el mismo análisis empezando a $t_f = 6\ de\ junio$. Si hay regiones con errores iguales o superiores al 10 %, se vuelve a repetir para $t_f = 5\ de\ junio$, y así sucesivamente. Finalmente, se mira cuántos días hacia atrás hemos tenido que ir para cada región, con tal de obtener un error que esté, de forma consistente, por debajo del 10 %. La siguiente tabla muestra los resultados de comparar las dos series en estudio, es decir, error cometido en la serie del 11 de junio con respecto a la actualización del 16 de junio.

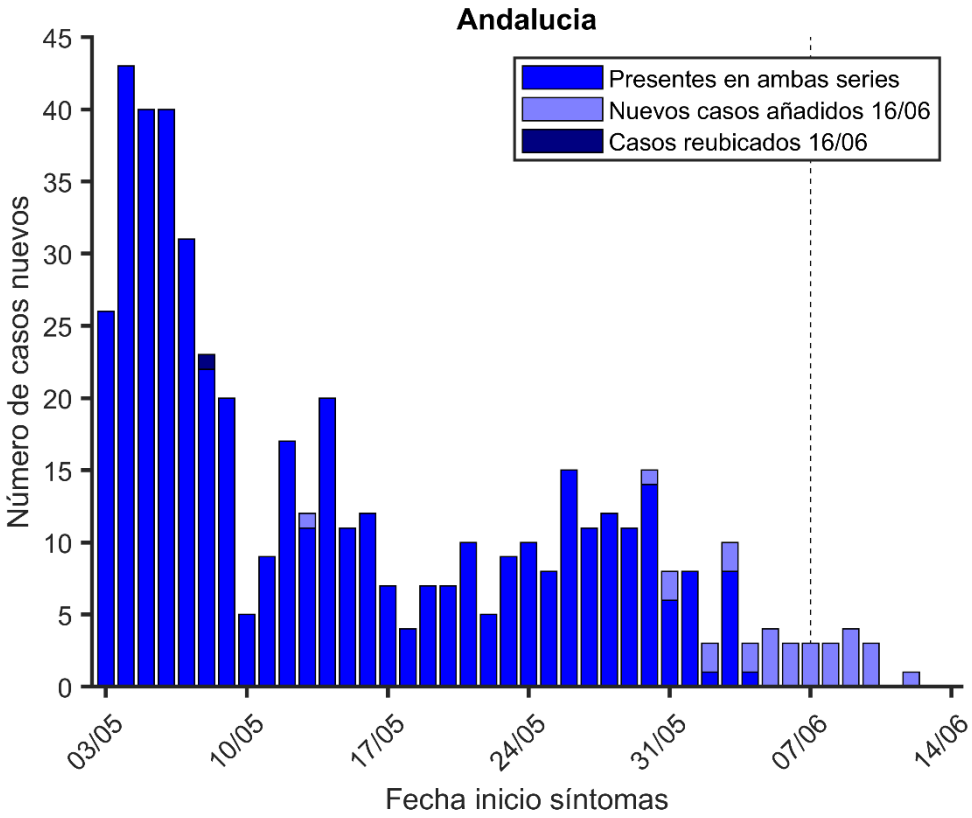
| | Días no fiables (error acum. ≥ 10%) |
|---|---|
| **Andalucía** | >7 |
| **Aragón** | 6 |
| **Asturias** | >7 |
| **Baleares** | 6 |
| **Comunitat Valenciana** | 5 |
| **Canarias** | >7 |
| **Cantabria** | 3 |
| **Castilla y León** | 5 |
| **Castilla-La Mancha** | >7 |
| **Catalunya** | 4 |
| **Ceuta** | NA |
| **Extremadura** | >7 |
| **Galicia** | 5 |
| **La Rioja** | NA |
| **Madrid** | >7 |
| **Melilla** | NA |
| **Murcia** | 6 |
| **Navarra** | 4 |
| **Euskadi** | 6 |
| **España** | 7 |

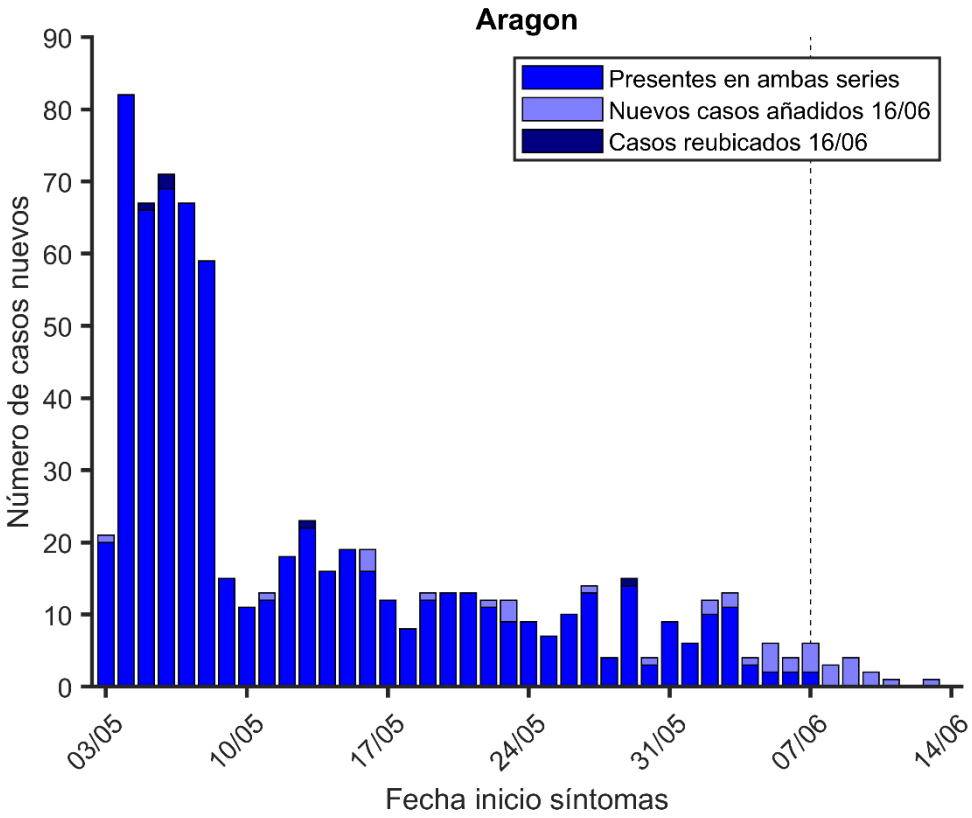
Estos resultados deben considerarse provisionales, ya que surgen de la comparación de sólo dos series y, además, están afectados no sólo por el retraso de los últimos días sino también por la revisión histórica (especialmente aquellas con resultado >7 días). Por otro lado, para aquellas comunidades con una incidencia muy baja, el error puede verse afectado por el ruido propio de estas situaciones. Los próximos días iremos actualizando estas cifras para poder corregirlas.

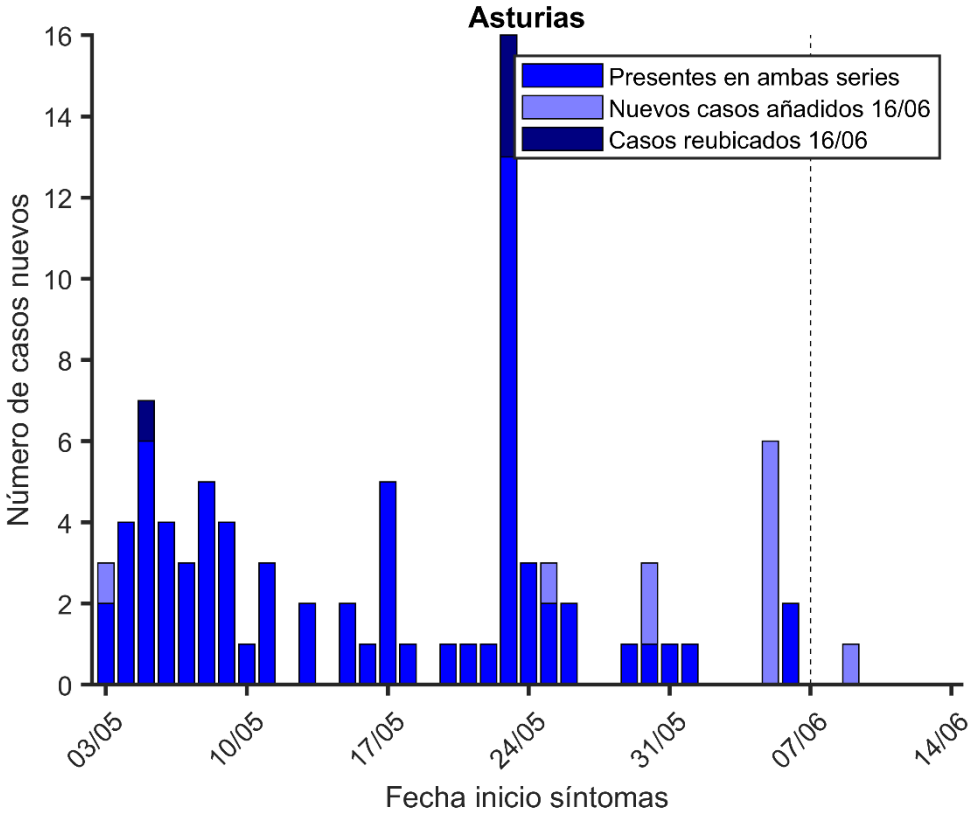
## Conclusiones

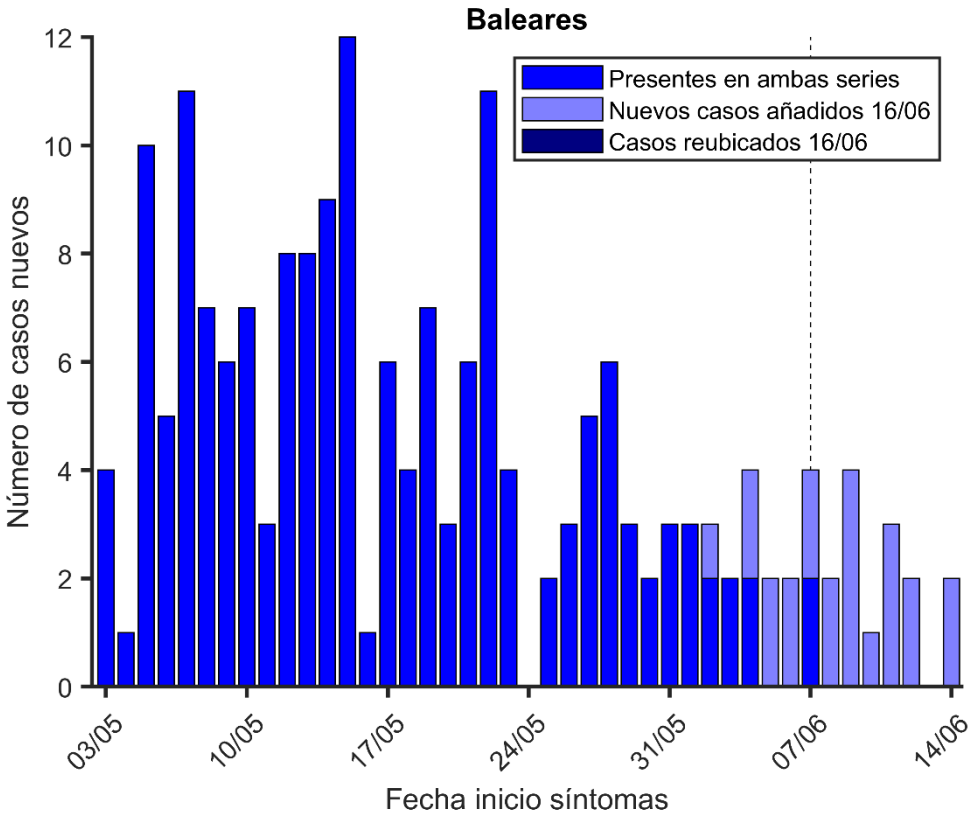
De este análisis se desprende que, efectivamente, hay un período de entre 3 y 7 días, en función de la comunidad autónoma, en el cual los datos deberían tomarse con cautela, ya que las cifras estarían subestimadas y serían corregidas al alza en actualizaciones posteriores. No obstante, el factor de la revisión de la serie histórica aún enmascara el efecto del retraso de los últimos días. **Este análisis deberá repetirse en días sucesivos para poder identificar claramente el periodo de retraso en el diagnóstico y notificación de los casos de los últimos días.** Una vez la serie histórica esté consolidada y presente pocas variaciones, la comparación de series sucesivas deberá dar la clave para estimar dicho retraso. De momento, los resultados parecen indicar que **los datos de los últimos 5-7 días no deberían ser tenidos en cuenta a la hora de analizar la situación actual**. Aunque puede parecer que es un intervalo grande, no hay que perder de vista que se está trabajando con fecha de inicio de síntomas, que de forma natural es unos días anterior a la fecha de registro. En este sentido, la mayoría de fuentes de datos oficiales trabajan con fecha de diagnóstico o con fecha de notificación, de manera que a la hora de comparar hay que tener en cuenta el retraso natural entre los tres tipos de registros.

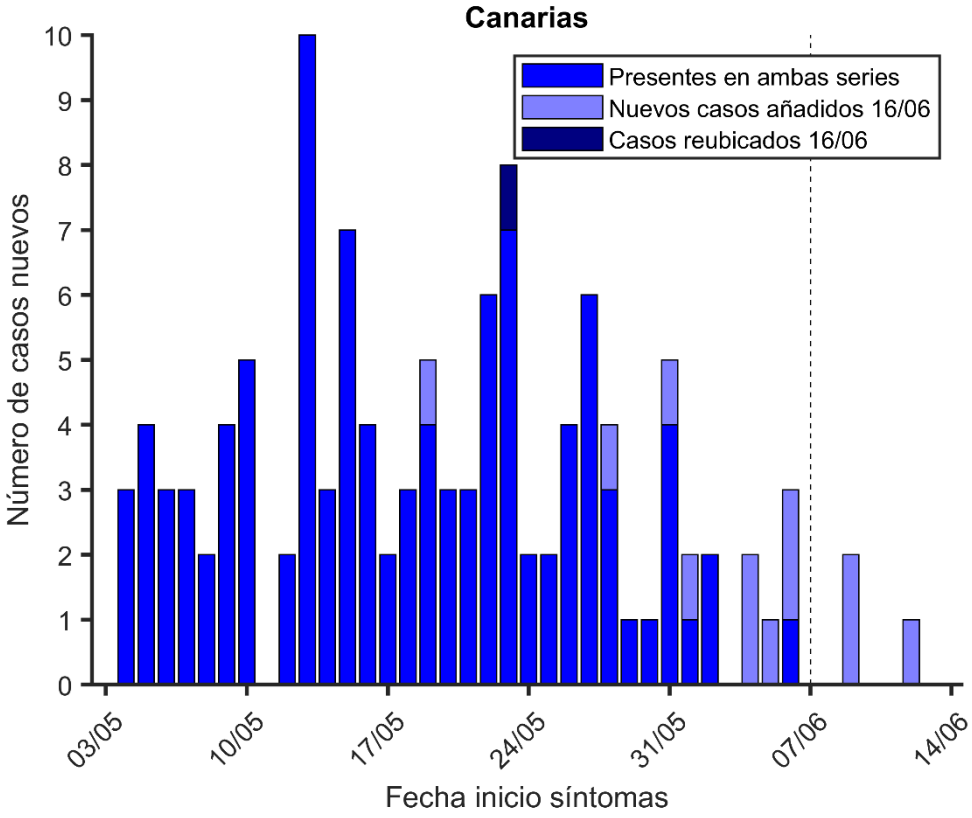
## Apéndice A. Comparación de las dos series para cada comunidad autónoma

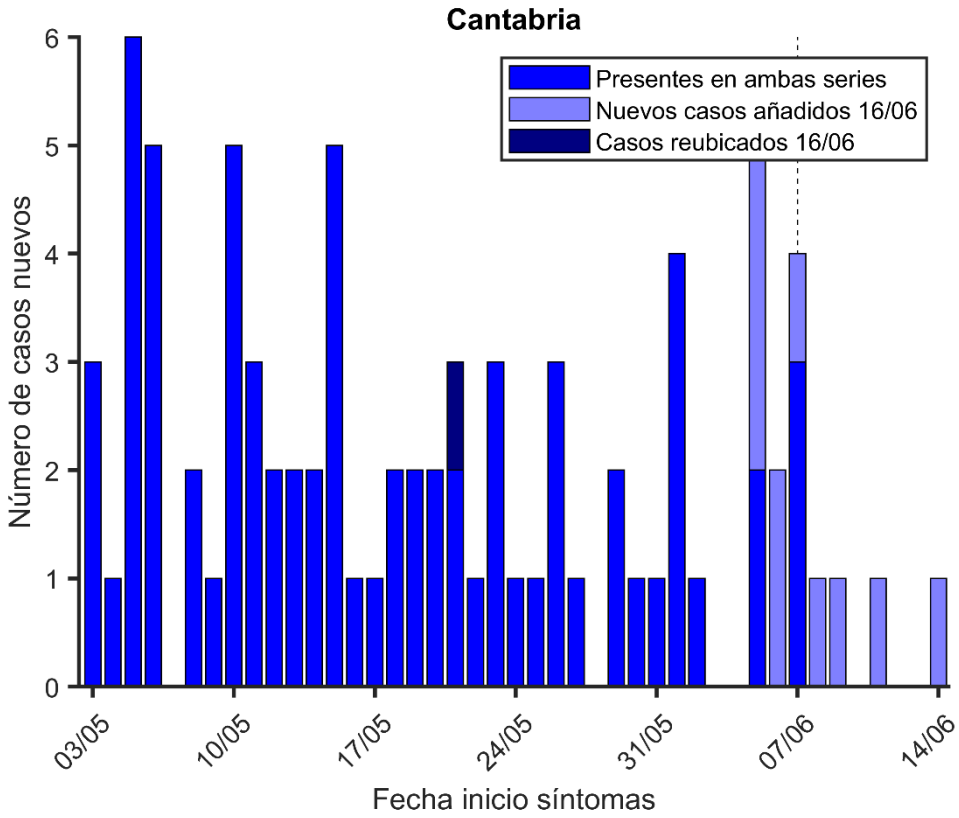


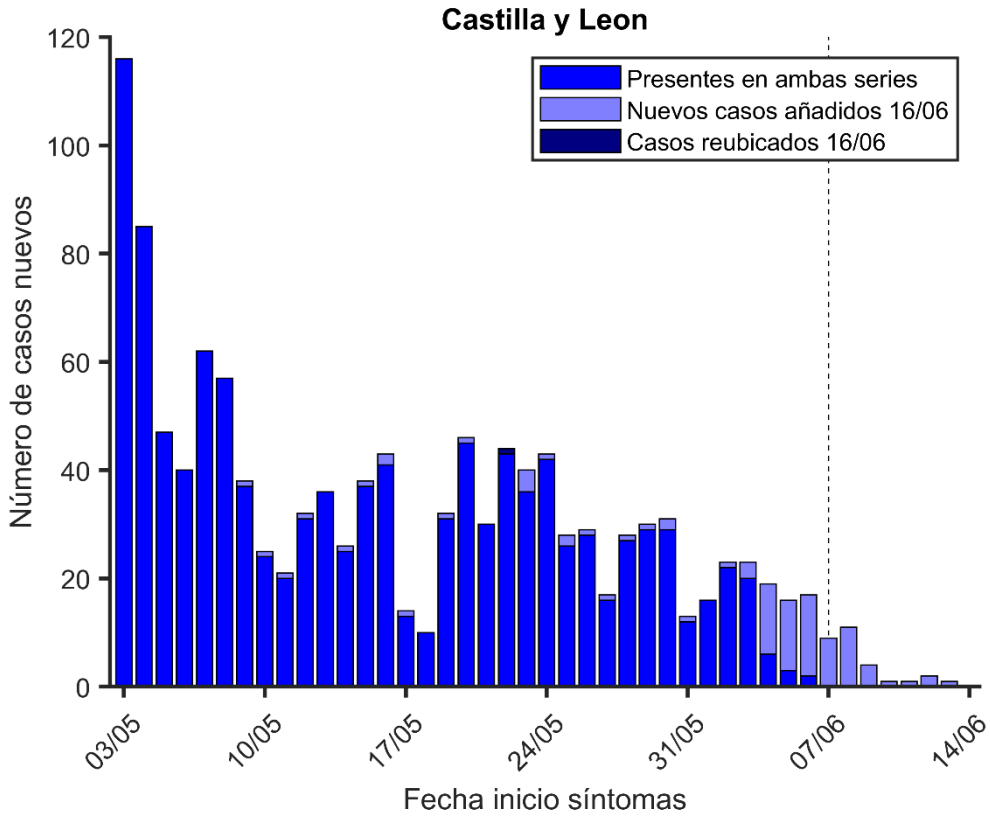



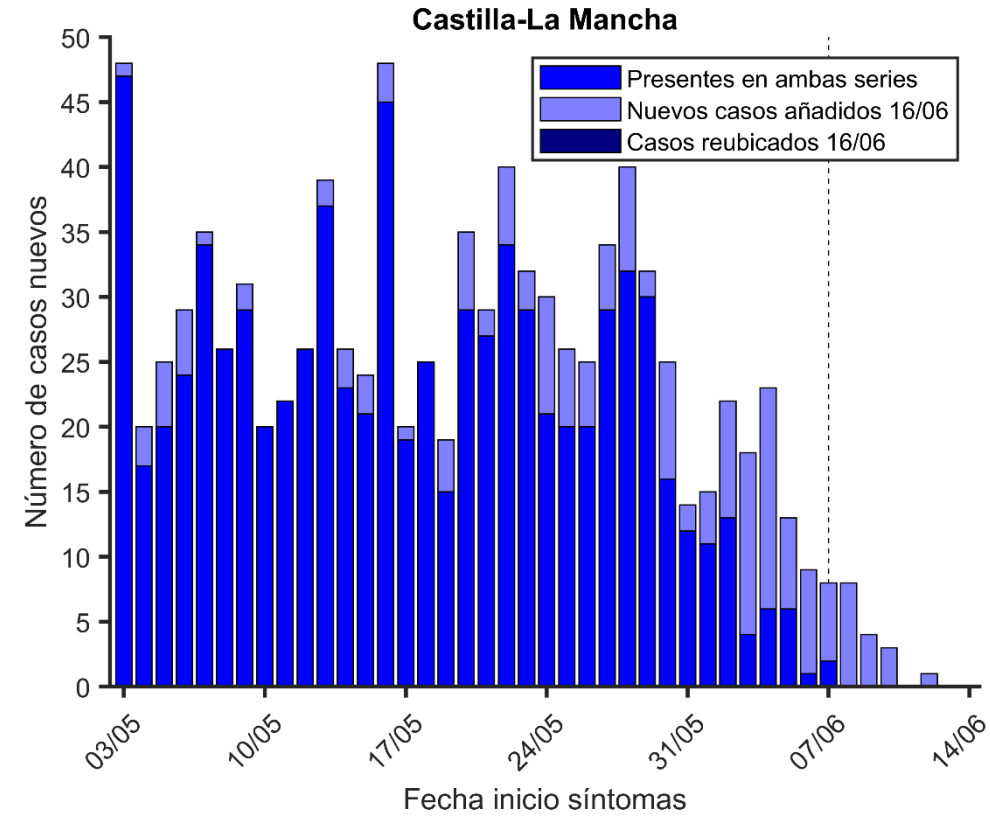

Asturias
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
2
4
6
8
10
12
14
16
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

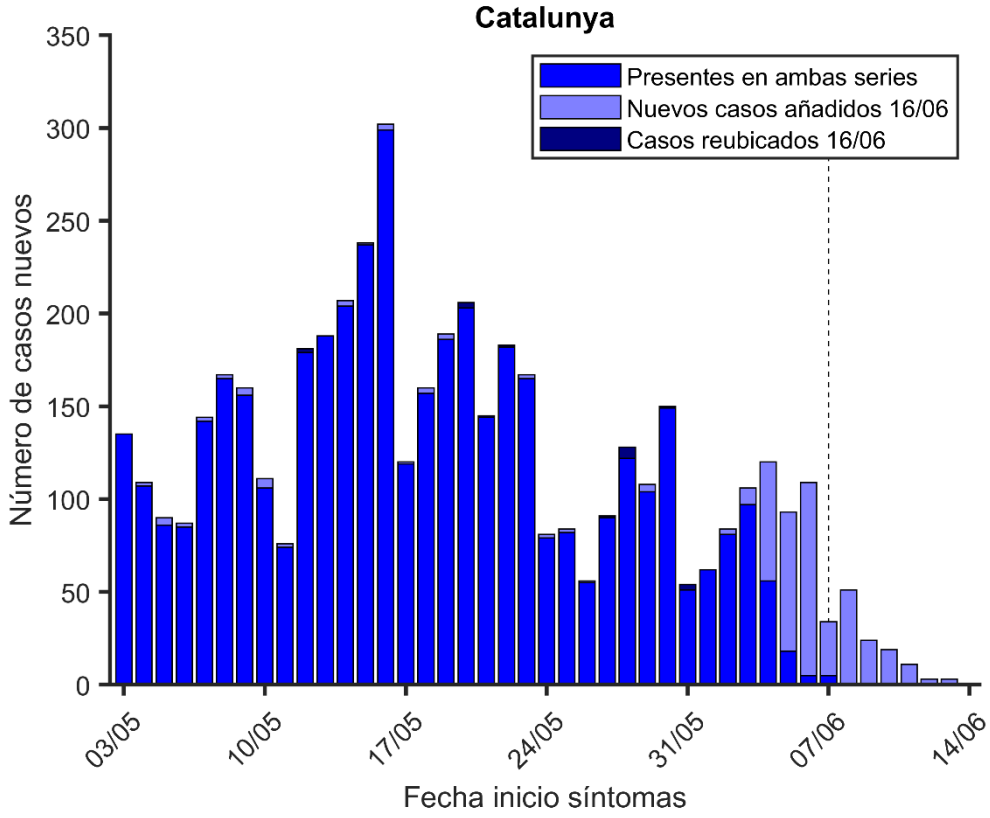

Baleares
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
2
4
6
8
10
12
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

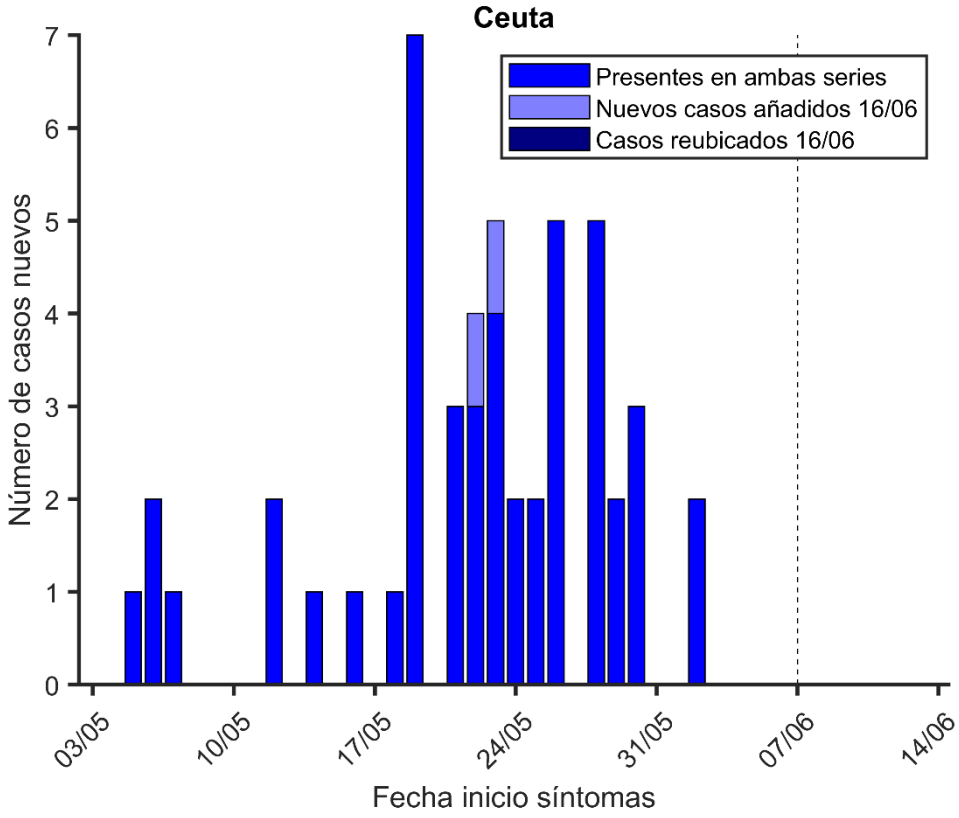

Canarias
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
1
2
3
4
5
6
7
8
9
10
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

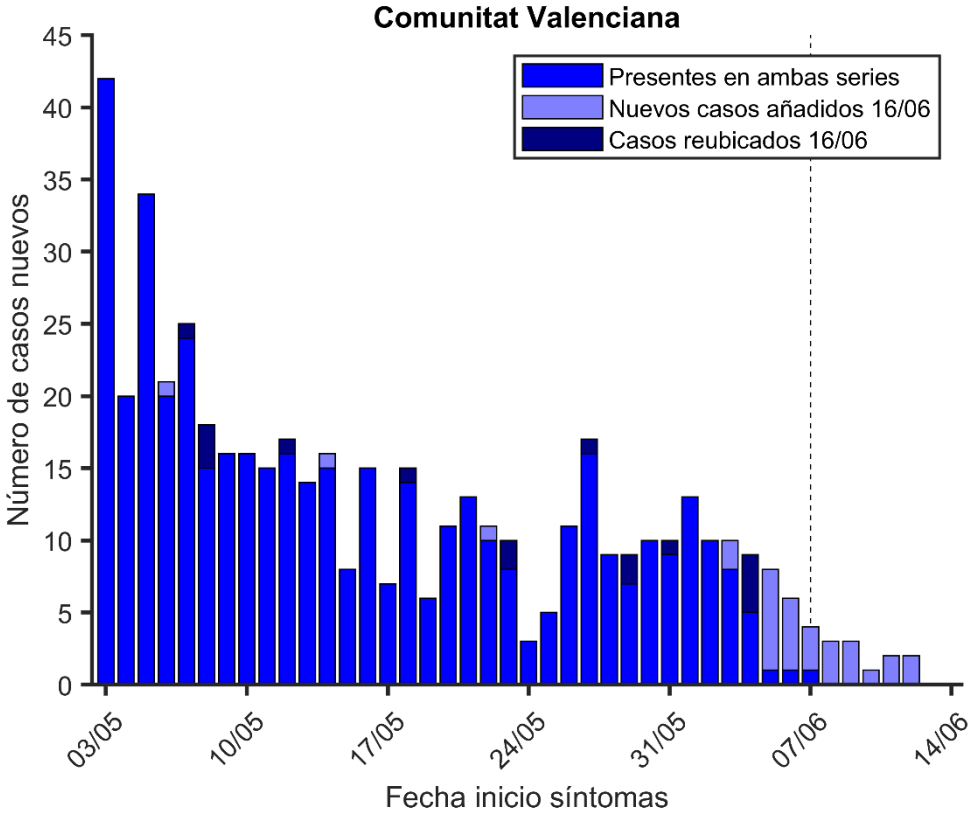

Cantabria
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
1
2
3
4
5
6
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

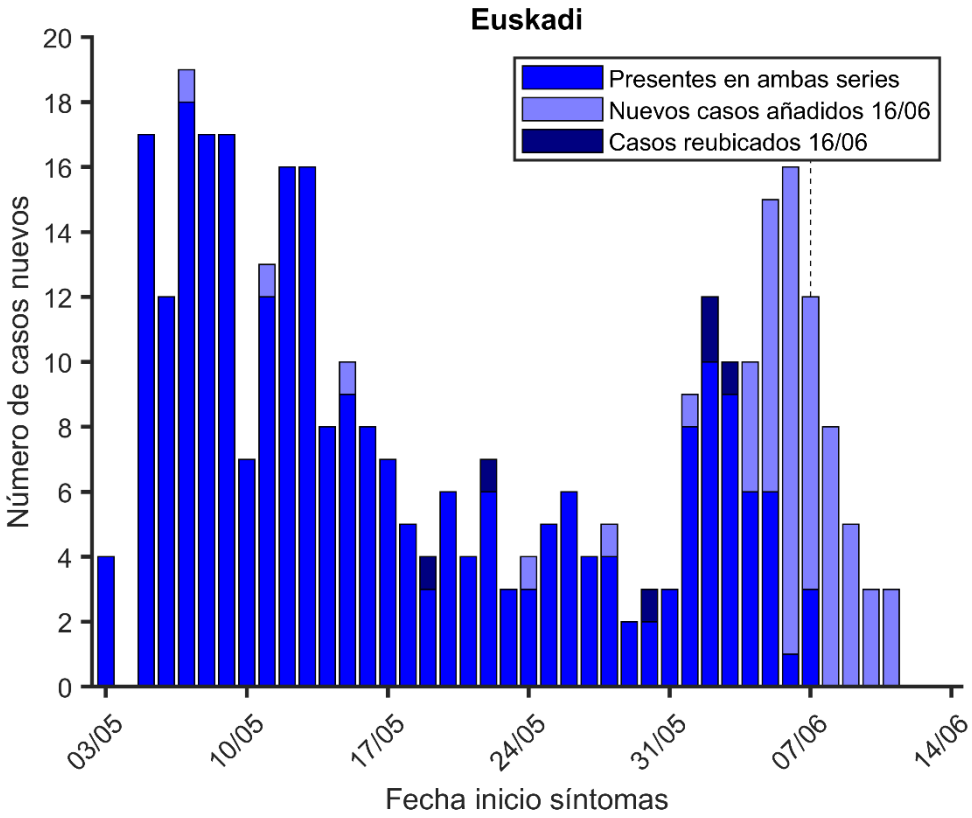

Castilla y Leon
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
20
40
60
80
100
120
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

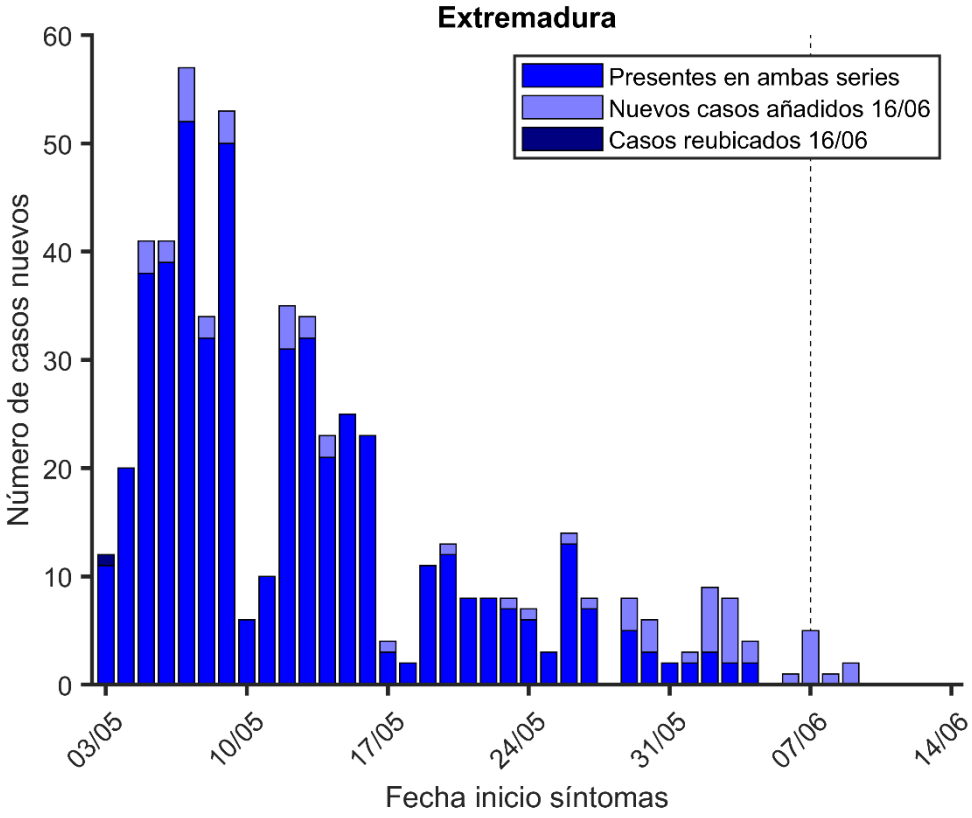

Castilla-La Mancha
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
5
10
15
20
25
30
35
40
45
50
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

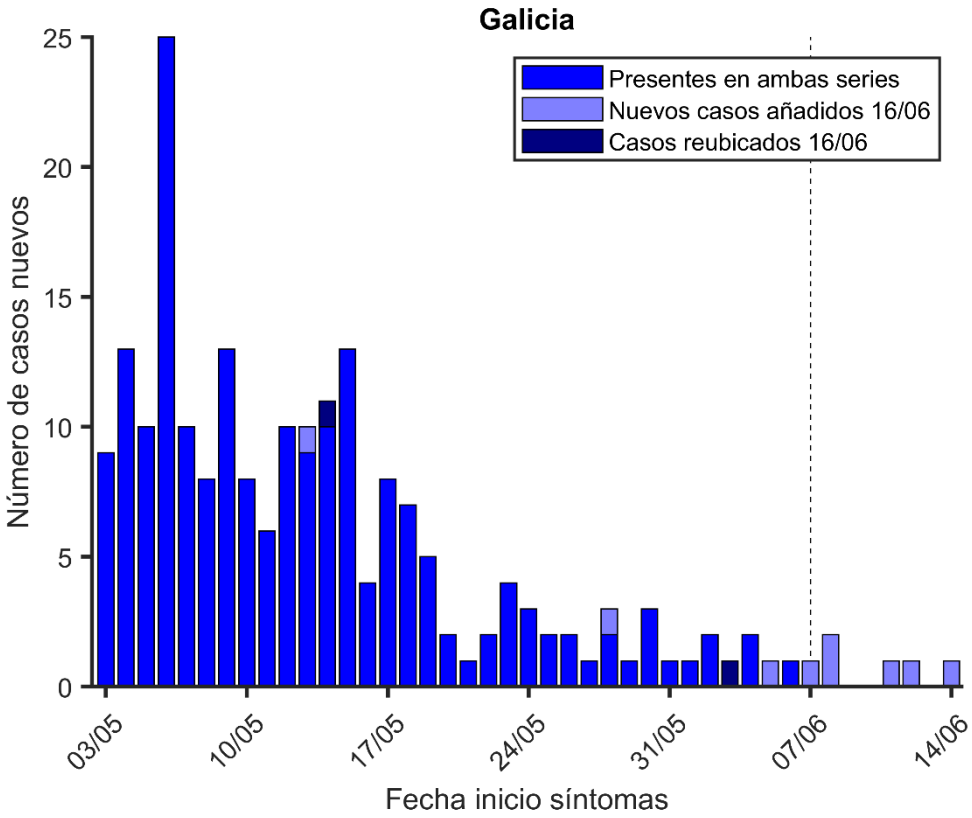

Catalunya
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
50
100
150
200
250
300
350
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

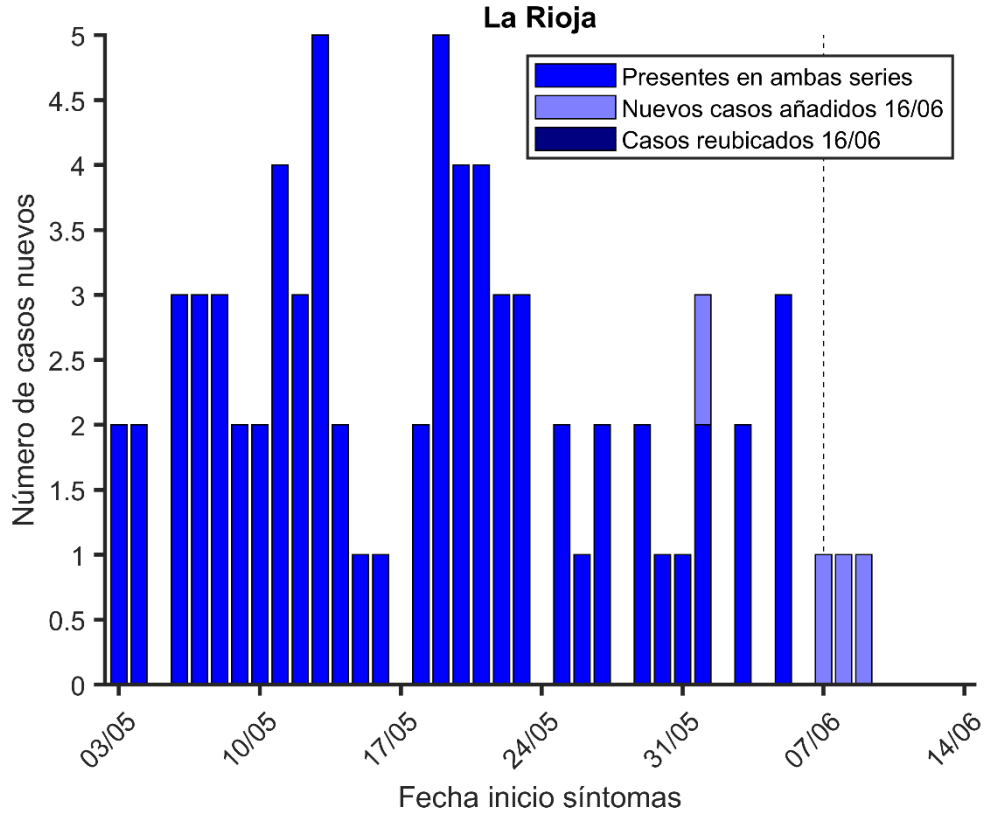

Ceuta
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
1
2
3
4
5
6
7
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

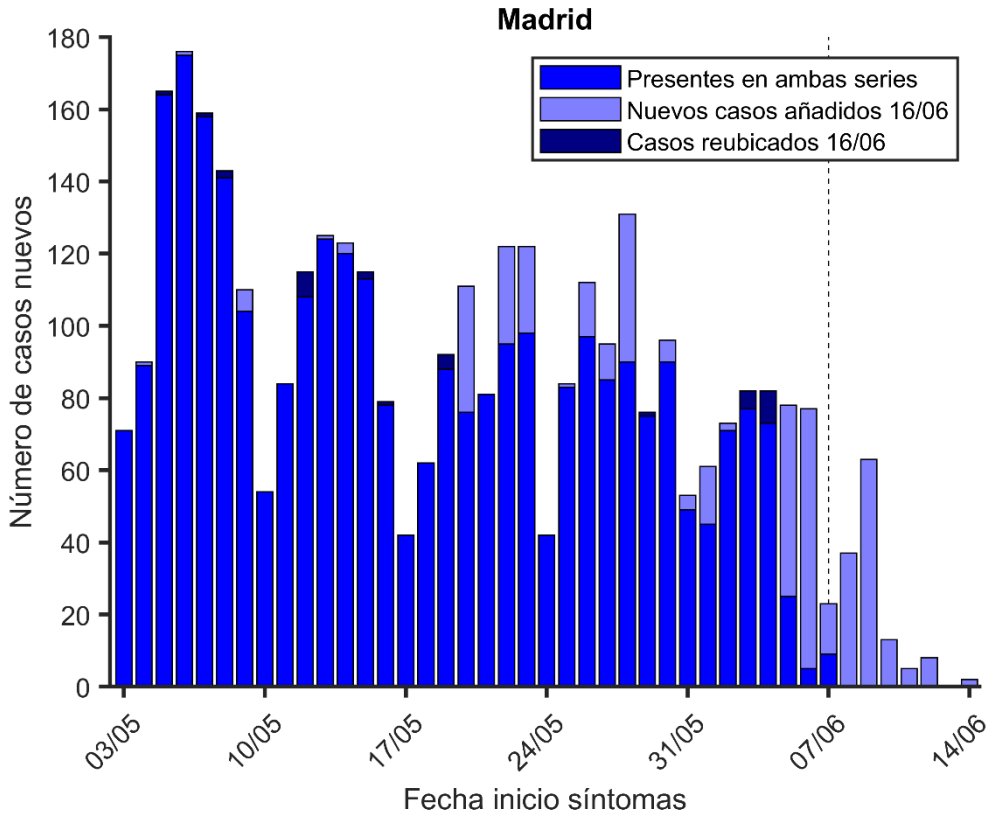

Comunitat Valenciana
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
5
10
15
20
25
30
35
40
45
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

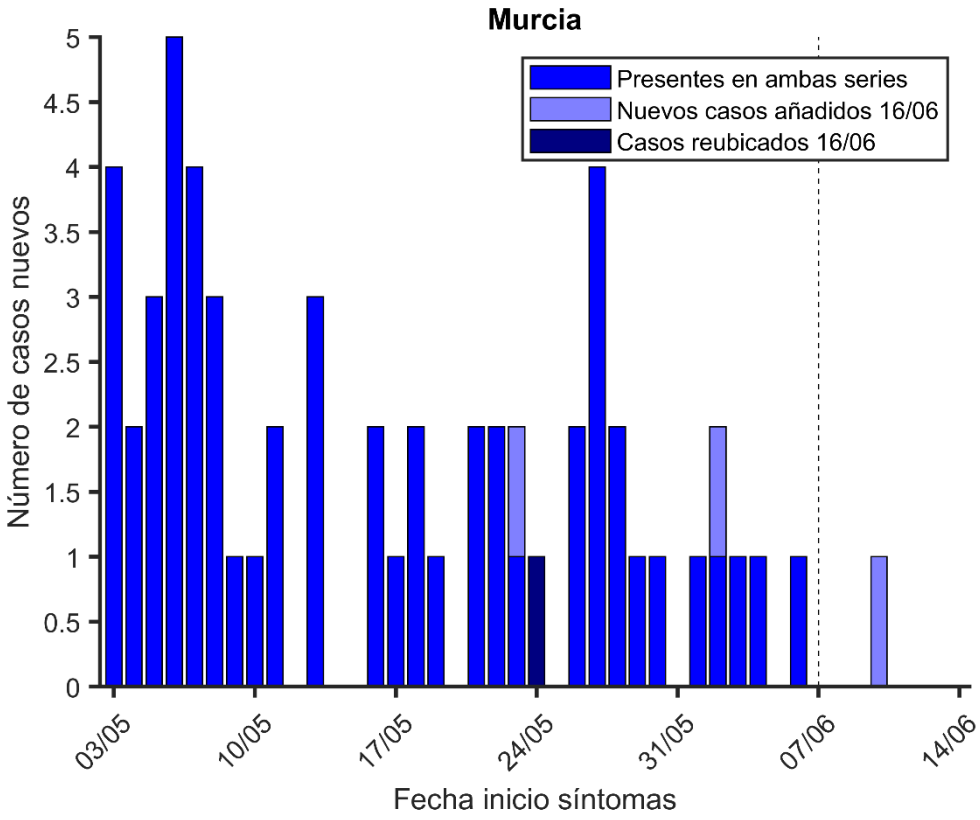

Euskadi
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
2
4
6
8
10
12
14
16
18
20
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas

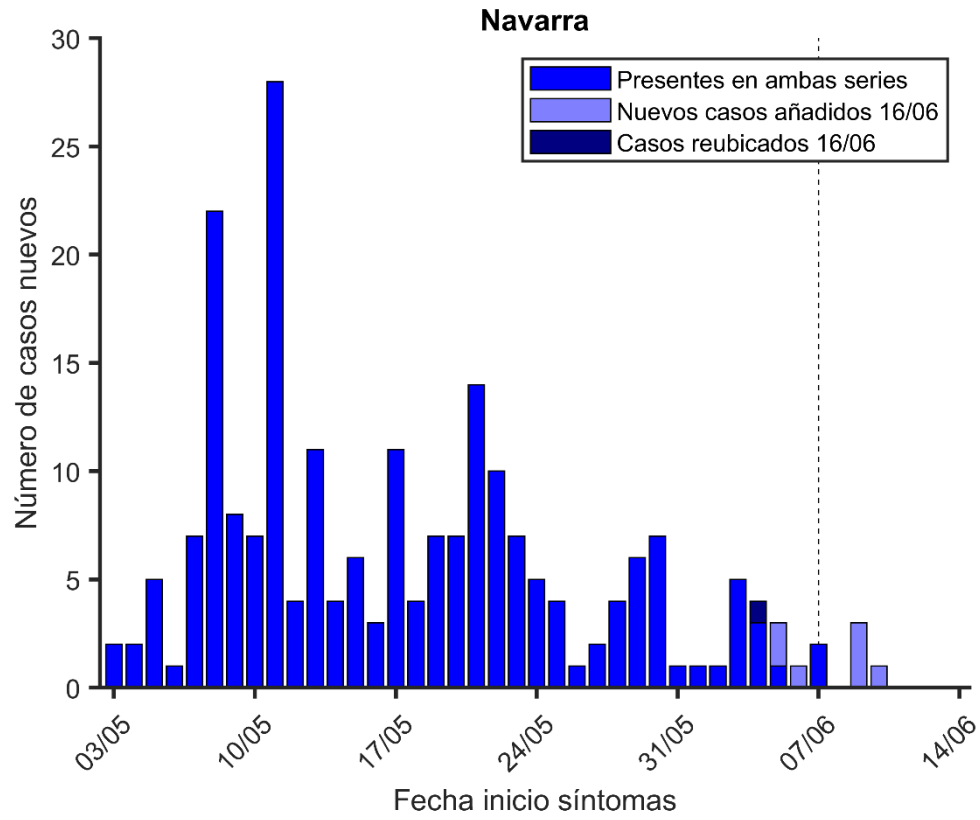

Extremadura
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
10
20
30
40
50
60
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas


Galicia
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
5
10
15
20
25
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas


La Rioja
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas


Madrid
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
20
40
60
80
100
120
140
160
180
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas


Murcia
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas


Navarra
Presentes en ambas series
Nuevos casos añadidos 16/06
Casos reubicados 16/06
Número de casos nuevos
0
5
10
15
20
25
30
03/05
10/05
17/05
24/05
31/05
07/06
14/06
Fecha inicio síntomas