

Análisis de la evolución de la incidencia de la COVID-19 en España desde el 1 de enero de 2020 hasta la actualidad

Versión abreviada para seguimiento de casos y mortalidad mediante R y R Markdown (RStudio ®)

Juan Matorras Díaz-Caneja

23/02/2021

1.- Introducción

Esta es una versión abreviada de **AnalisisCovidEspana.Rmd** donde se analizan con más detalle que aquí los datos de la incidencia de la COVID-19 en España desde el comienzo del año 2020.

La razón de ser de esta versión es alejarse del ejercicio académico que representaba el documento original para condensar su contenido y centrarse en los datos más relevantes para el seguimiento de la incidencia de la enfermedad y de su mortalidad asociada.

Por su propia esencia, éste es un documento vivo que además de ser puesto al día periódicamente con los nuevos datos disponibles, va sufriendo adiciones, modificaciones y correcciones de erratas. La última versión disponible de este documento y de los datos empleados en su elaboración se pueden encontrar en el repositorio de GitHub: <https://github.com/JuanMatorras/Covid-19>.

Los datos de partida son los publicados por el Gobierno de España a través de diferentes organismos públicos. Para información detallada sobre el origen de los datos, consúltase el documento principal mencionado del que éste deriva.

2.- Proceso metodológico y software utilizado

El archivo de datos no ha sido sometido a ningún tipo de modificación o alteración previa y su manipulación en este análisis es el mínimo imprescindible para permitir el tratamiento de los datos y obtención de resultados.

Al ejecutar el código se descargan los archivos de datos directamente de la web si no se encuentran ya disponibles en el directorio /data.

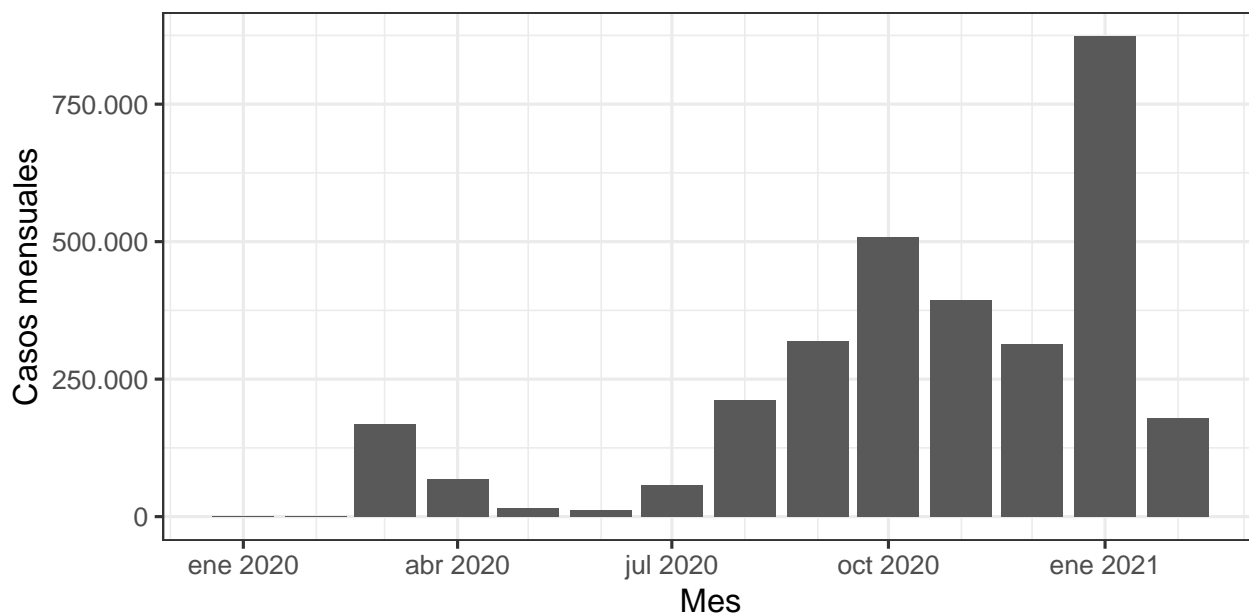
La fecha y hora de descarga de los datos que han sido utilizados para la elaboración de las tablas y gráficos incluidos en este informe ha sido (aaaa-mm-dd hh:mm:ss): **2021-02-23 10:01:42**

El análisis se ha llevado a cabo utilizando el entorno de desarrollo integrado de **RStudio** © versión 1.4.1103 (1) para el software libre de análisis estadístico **R**, versión 4.0.3 (2), complementado con los paquetes: **lubridate** ver.1.7.9.2 (3), **knitr** ver.1.31 (4), **tidyverse** ver.1.3.0 (5), **ggplot2** ver.3.3.3 (6) y **data.table** ver.1.13.6 (7).

Nota: Hay que llamar la atención sobre la naturaleza de los datos, los cuales proceden de la agregación de las aportaciones desde las CCAA al sistema RENAVE. Esto resulta en que los datos de casos de fechas recientes no están completos y esto hace que todas las gráficas con desagregación diaria o variables derivadas, como la incidencia acumulada, presenten siempre un tramo descendente en su extremo final, con independencia de si realmente el progreso de la enfermedad es de expansión o de contracción.

3.- Casos por meses y número total de casos detectados desde el inicio de 2020

La evolución de número de casos notificados por meses se refleja en el gráfico que se muestra a continuación:



Correspondiente a los valores que se incluyen en la tabla siguiente:

Mes	Casos
ene 2020	45
feb 2020	1.728
mar 2020	167.358
abr 2020	68.311
may 2020	16.100
jun 2020	11.134
jul 2020	57.608
ago 2020	212.231
sep 2020	318.939
oct 2020	508.200
nov 2020	394.013
dic 2020	313.960
ene 2021	873.169
feb 2021	178.907

El número total de casos acumulados desde el 1 de enero de 2020 hasta la fecha indicada en el punto anterior según los datos oficiales disponibles en ese momento ascienden a un total de **3.121.703** personas.

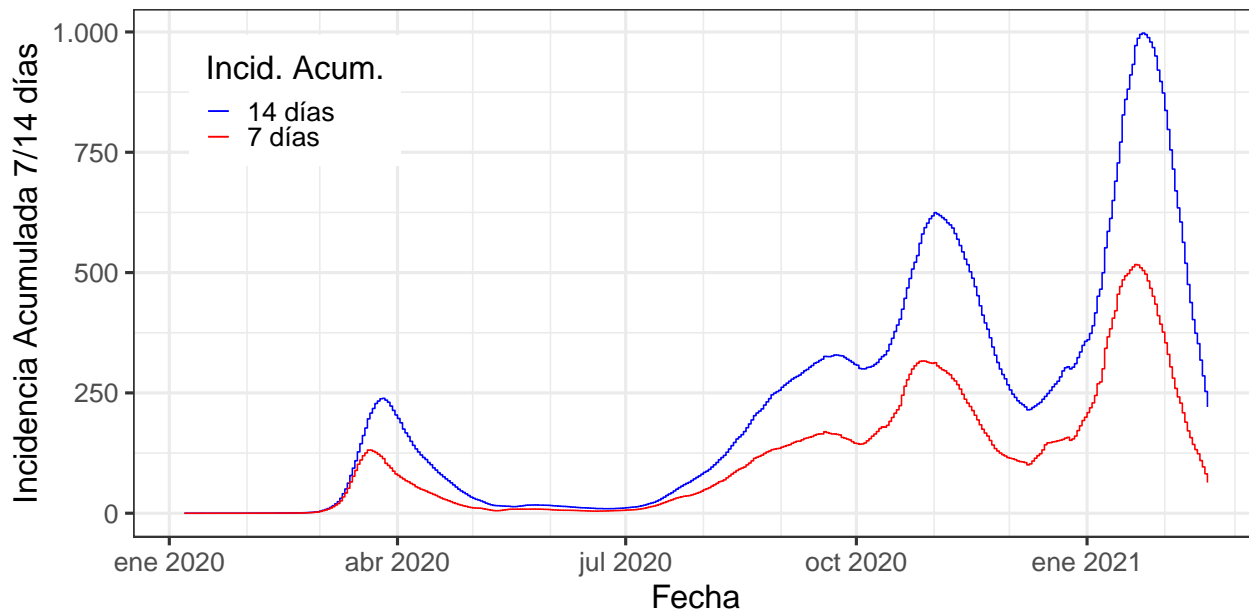
Nota: Durante el mes de enero de 2021 aparecieron datos discrepantes correspondientes a enero de 2020 por errores en el registro de nuevos casos que se habían asignado incorrectamente al año 2020 cuando correspondían al 2021. En la segunda semana de febrero esta deficiencia parece haber sido solucionada ya que el número total de casos de ene-2020 han vuelto a los valores que se reflejaban antes del cambio de año.

Considerando una población en España de **47,332** millones de personas según los datos publicados por el INE (Instituto Nacional de Estadística) correspondientes al inicio del año 2020, el porcentaje de contagio de la población es del **6,595 %** hasta la fecha. Insistimos en que, puesto que la incidencia de la enfermedad en los primeros meses de 2020 está por fuerza minusvalorada por la escasez de pruebas de diagnóstico, y no pudiendo olvidar que además tenemos el fenómeno de los casos de infección asintomáticos, el porcentaje de población afectada realmente es necesariamente más alto que el reflejado en este cálculo y es más que probable que sea del orden del doble. De hecho, para poder cubrir esta laguna y tener una idea más fiable del verdadero impacto de la enfermedad sobre el total de la población están los estudios de sero-epidemiología

que se han venido realizando desde junio de 2020 (<https://portalcne.isciii.es/enecovid19/>).

4.- Evolución de la incidencia acumulada en 14 días (IA14) y 7 días (IA7)

En el siguiente gráfico se representan las incidencias acumuladas por cada 100.000 habitantes correspondientes a periodos de 14 y 7 días:

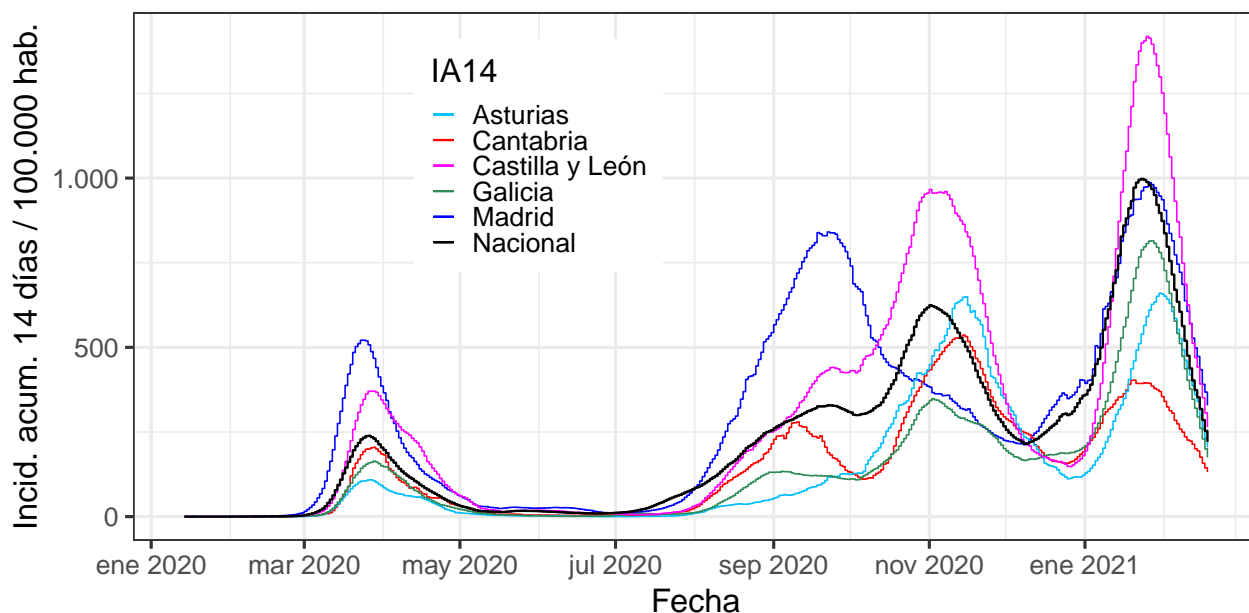


Obviando el hecho ya comentado de la subestimación del primer pico de la ola por la baja realización de pruebas de diagnóstico, queda claro que, a pesar de que los medios de comunicación y, en consecuencia, la propia población general hablen de una tercera ola de la pandemia que arranca en el mes de diciembre, no habiendo bajado la IA14 de 200 sería más apropiado hablar de una segunda fase de la segunda ola puesto que ésta no se puede decir que llegó a estar bajo control. Si analizamos el impacto de la enfermedad desde el punto de vista de la mortalidad (ver apartado 9) sí que podríamos afirmar de algún modo que la segunda ola habría terminado a finales de diciembre, pero finalmente se ha fundido con el periodo de exceso posterior.

5.- Comparación de la IA14 por CCAA y nacional

En el siguiente gráfico se compara la evolución de la enfermedad entre comunidades muy diferentes (Comunidad Autónoma de Madrid, Cantabria, Asturias, Galicia, Castilla y León), junto con la evolución a nivel nacional en términos de incidencia aculada en 14 días por cada 100.000 habitantes.

Los datos de población en cada comunidad para el cálculo son disponibles en el momento en el INE (<https://www.ine.es/dynInfo/Infografia/Territoriales/capitulo.html#!tabla>).



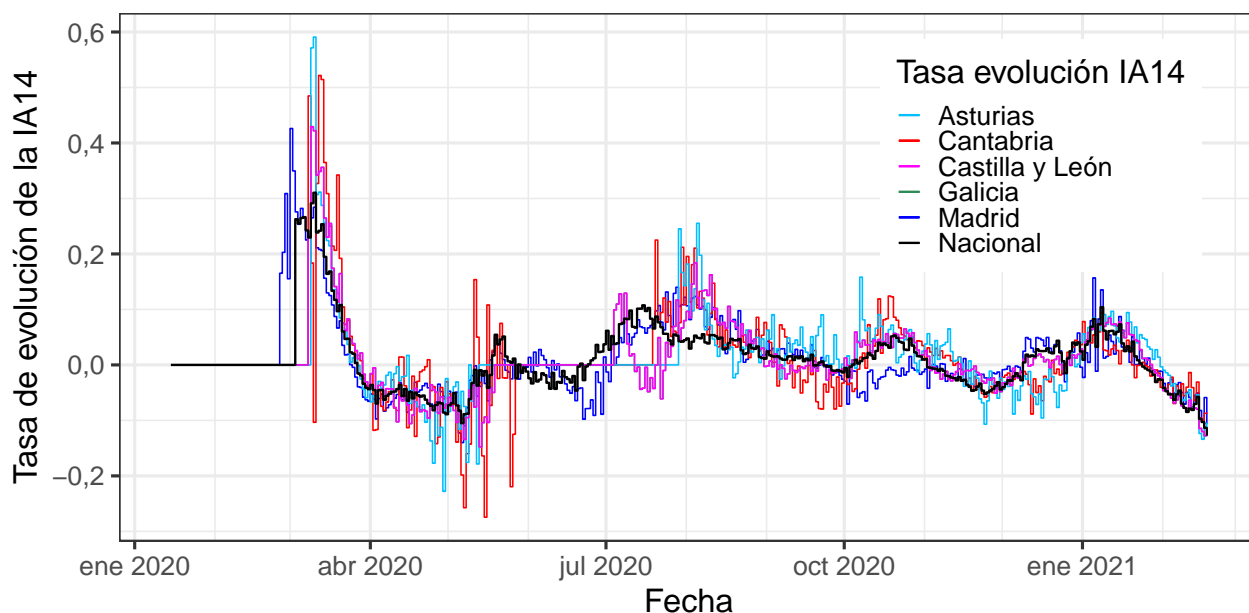
6.- Tasa de variación de la incidencia acumulada en 14 días (IA14)

En el siguiente gráfico se representa la evolución de la variación de la incidencia acumulada en 14 días por cada 100.000 habitantes (IA14) expresando la tasa de variación de esta incidencia como variación porcentual de la misma:

$$\text{tasaIA14}(i) = (\text{IA14}(i) - \text{IA14}(i-1)) / \text{IA14}(i-1)$$

Se impone la condición para el cálculo de la tasa que $\text{IA14}(i-1)$ sea mayor que cero para evitar obtener tasas de crecimiento infinitas por la división con denominador cero y la indeterminación $0/0$ en los casos de secuencias de IA14 con valor 0 en los inicios de las series temporales.

Como para el caso de incidencias acumuladas de valores muy bajos, pequeños cambios de la dicha incidencia representan cambios de tasa de evolución muy elevadas al ser el divisor pequeño, modificaremos la condición indicada en el párrafo anterior, exigiendo que la incidencia acumulada en el día anterior tenga como mínimo un valor de 5 casos por cada 100.000 habitantes.



Hay que hacer notar que, aunque la gráfica resultante tenga una apariencia similar a la del Número reproductivo básico instantáneo - R_t (número promedio de casos secundarios que cada sujeto infectado puede llegar a infectar en una etapa de tiempo (t)), no se trata de este indicador, cuyo cálculo es totalmente diferente al presentado aquí. El número reproductivo básico instantáneo calculado por el Instituto de Salud Carlos III puede ser encontrado en el siguiente enlace: <https://cnecovid.isciii.es/covid19/#ccaa>. Nótese que el nivel de referencia del número reproductivo es 1 mientras que para la tasa de evolución de la IA14 es 0.

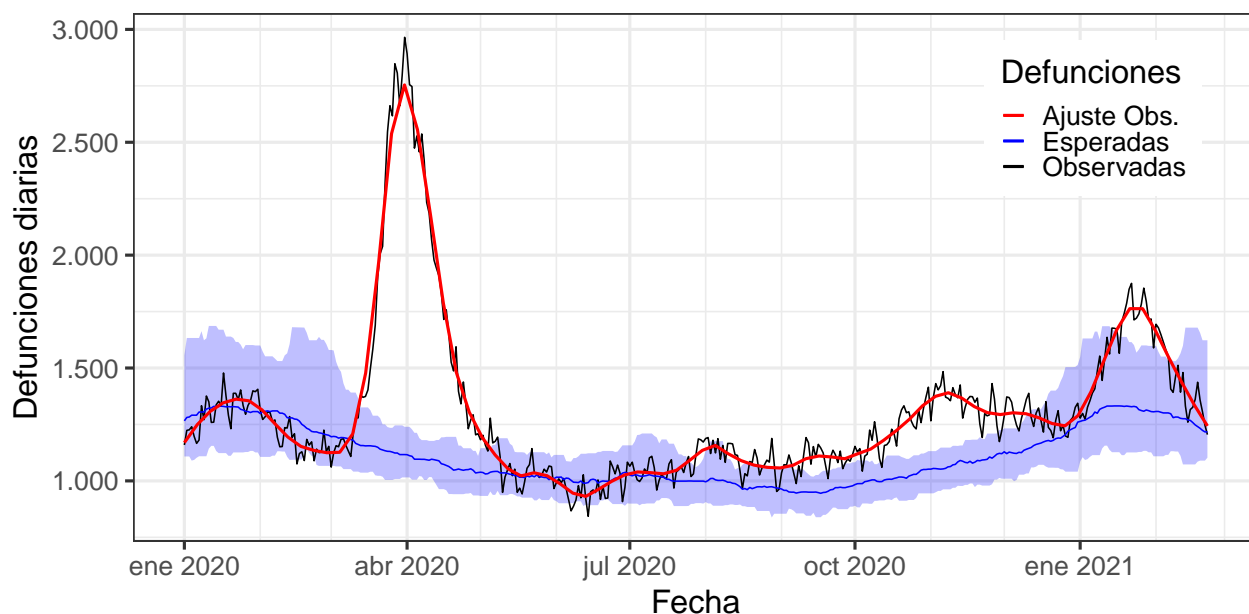
Por otro lado es lógica esta similitud entre las gráficas de la tasa de evolución de la IA14 y la del número reproductivo básico instantáneo, ya que números reproductivos altos se corresponden con evoluciones crecientes en la incidencia de la enfermedad mientras que números reproductivos por debajo de 1 marcan evoluciones decrecientes de la incidencia en el número de casos.

7.- Datos de mortalidad

Los datos se han obtenido del enlace del **Instituto de Salud Carlos III**: <https://momo.isciii.es/public/momo/data>.

La fecha y hora de descarga de los datos de mortalidad utilizados para la elaboración de los siguientes gráficos y tablas fue (aaaa-mm-dd hh:mm:ss): **2021-02-23 10:01:46**

Representemos en primer lugar la evolución del número de defunciones en comparación con las esperadas y su rango para los percentiles 1 y 99:



Se ha agregado una línea de ajuste estadístico de las defunciones observadas para facilitar la visualización de la evolución de las mismas suavizando los dientes de sierra propios de las observaciones por tratarse de datos diarios.

8.- Exceso de mortalidad

Técnicamente se define “periodo de exceso de mortalidad” cuando se cumplen las siguientes condiciones:

- Se observa al menos dos días consecutivos con defunciones observadas por encima del percentil 99 de las estimadas.
- La fecha de inicio del periodo es el primer día con las defunciones observadas por encima de las estimadas.
- La fecha de fin del periodo es el último día con las defunciones observadas por encima de las estimadas.

- Si entre la fecha de fin de un periodo y la fecha de inicio del siguiente hay dos días, se unifican ambos periodos, tomando la fecha de inicio del primer periodo y fecha de fin del último.

Con estas premisas podemos aislar los periodos en los que se han producido dichas circunstancias y calcular el exceso de defunciones durante esos lapsos de tiempo concretos.

Antes de pasar a ello hagamos una comparación directa de las cifras acumuladas hasta la fecha. En este caso tenemos un total de **547.127** defunciones observadas y **466.676** defunciones esperadas, resultando un exceso de **80.451** defunciones. Expresando dicho exceso en términos porcentuales, nos encontramos con un **17,2 %** más fallecimientos de los esperados.

Retomando la senda de la ortodoxia y aplicando ahora sí los criterios técnicos “oficiales” que presentábamos más arriba que definen los periodos de exceso de mortalidad, las fechas que delimitan el principio y final de los periodos de exceso padecidos a lo largo de 2020 son (fechas en formato aaaa-mm-dd):

- Antes de unificar periodos de exceso próximos:

Inicio	Fin
2020-03-10	2020-05-09
2020-07-20	2020-08-29
2020-09-01	2020-12-25
2020-12-28	2021-02-13

- Después de unificar los periodos de exceso cercanos (adyacentes o con un día intermedio):

Inicio	Fin
2020-03-10	2020-05-09
2020-07-20	2020-08-29
2020-09-01	2020-12-25
2020-12-28	2021-02-13

Los excesos de defunciones en estos 4 periodos son:

Inicio	Fin	Exceso de defunciones
2020-03-10	2020-05-09	44.550
2020-07-20	2020-08-29	4.673
2020-09-01	2020-12-25	21.489
2020-12-28	2021-02-13	12.256

Siendo el total agregado de exceso de defunciones de **82.968** personas.

Los excesos en términos porcentuales con respecto a las esperadas, pero circunscritos exclusivamente al propio periodo de exceso de defunciones y dejando fuera el resto de la serie temporal:

Inicio	Fin	Exceso de defunciones	Porcentaje de exceso
2020-03-10	2020-05-09	44.550	66,8
2020-07-20	2020-08-29	4.673	11,5
2020-09-01	2020-12-25	21.489	17,7
2020-12-28	2021-02-13	12.256	19,6

Aunque en el exceso de defunciones haya casos de fallecimiento no directamente imputables a la COVID-19, hay

que asignar dichas muertes a la crisis del COVID-19. Si determinadas patologías no son debidamente atendidas en tiempo y forma por la sobrecarga del sistema sanitario provocada por la pandemia, los fallecimientos asociados a las mismas son por tanto atribuibles a la COVID-19 aunque el virus no haya sido la causa directa del fallecimiento correspondiente.

El índice de mortalidad de la COVID-19 en España en el periodo de estudio, medido como exceso de mortalidad atribuible directa o indirectamente a la COVID por cada mil habitantes, es de **1,75**.

No podemos dejar de llamar la atención sobre el hecho de que en la determinación de las cifras de exceso de defunciones se ha utilizado como nivel de referencia el número de defunciones esperadas. Es perfectamente argumentable que durante el periodo de estado de alarma este nivel de comparación debería ser inferior al estadísticamente obtenido con datos de años previos ya que el propio estado de alarma tuvo por necesidad incidencia en el número de fallecimientos por accidente laboral y por accidente de tráfico, sin duda disminuyéndolos. Consecuentemente debería rebajarse el patrón de referencia de defunciones esperadas durante el estado de confinamiento y el exceso de defunciones por causa de la COVID-19 sería superior al mostrado más arriba. Aunque es posible realizar estimaciones de estas desviaciones con datos disponibles públicamente, dejamos esa posibilidad de perfeccionamiento del estudio para mejor oportunidad.

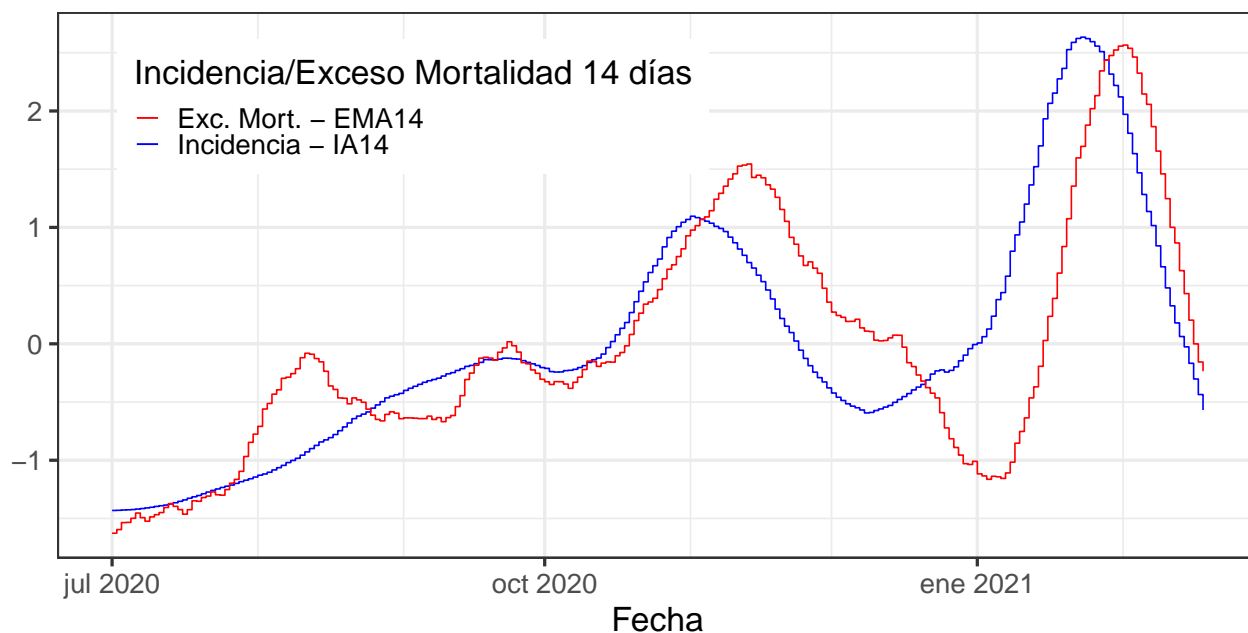
9.- Exceso de mortalidad acumulada en 14 días (EMA14) y comparación con la IA14

El gráfico que se incluye a continuación incluye la representación escalada (estadísticamente) de las dos variables siguientes para tener una idea visual de su correlación:

- Incidencia diaria acumulada en 14 días por cada cien mil (100.000) habitantes.
- Exceso de mortalidad acumulada en 14 días por cada millón (1.000.000) de habitantes.

Definimos exceso de mortalidad acumulada en 14 días por millón de habitantes como la diferencia entre la mortalidad diaria acumulada observada menos la mortalidad acumulada esperada, calculadas ambas en periodos acumulativos de 14 días.

Se representan datos correspondientes a fechas posteriores al 1 de julio de 2020 para evitar la distorsión que se genera durante la primera ola de la pandemia por la infravaloración del número de casos por la falta de disponibilidad de pruebas diagnósticas.



.....

Apéndice - Referencias

- (1) RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. URL: <http://www.rstudio.com/>
- (2) R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- (3) Garrett Golemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL: <http://www.jstatsoft.org/v40/i03/>
- (4) Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.
- (5) Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- (6) H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- (7) Matt Dowle and Arun Srinivasan (2020). data.table: Extension of `data.frame`. R package version 1.13.2. <https://CRAN.R-project.org/package=data.table>