

APUNTES - 2020/15

**Datos COVID-19: Actualizaciones, confusiones y estimaciones de
contagiados reales**

José E. Boscá (U. de Valencia y FEDEA)
Rafael Doménech (BBVA Research y U. de Valencia)
Javier Ferri (U. de Valencia y FEDEA)

Octubre 2020

fedea

Las opiniones recogidas en este documento son las de sus autores y no coinciden necesariamente con las de FEDEA.

Datos COVID-19: Actualizaciones, confusiones y estimaciones de contagiados reales.

José E. Boscá (U. de Valencia y FEDEA)
Rafael Doménech (BBVA Research y U. de Valencia)
Javier Ferri (U. de Valencia y FEDEA)

20 de octubre de 2020

Resumen

Este documento muestra cómo las depuraciones y actualizaciones oficiales de los casos diarios reportados de Covid-19 pueden alterar significativamente los valores calculados del Índice Acumulado por 100.000 casos. También se presenta evidencia de la importancia de controlar tanto por el número de pruebas diagnósticas realizadas, como por la composición de edades de los contagiados, para entender la brecha existente entre casos diagnosticados y hospitalizaciones entre la primera y la segunda olas del virus.

La polémica entre el gobierno central y el autonómico de Madrid sobre las medidas a adoptar para doblegar la curva de casos en la segunda oleada de la pandemia ha propiciado un agrio debate político, periodístico y social. En el centro del debate ha estado la incidencia de casos acumulados por cien mil habitantes (también otros dos indicadores de ocupación de UCIs y de positividad de las pruebas). En este post queremos hacer algunas reflexiones acerca de la calidad de los datos disponibles y si ésta permite realizar dichas discusiones de manera que conduzcan a las mejores decisiones por parte de los responsables de las administraciones públicas. Es un tema en el que ya insistimos (desde una perspectiva diferente) en un [post anterior](#).

Es obvio que una correcta medición del número de nuevos casos que se confirman cada día proporciona información muy relevante para entender la evolución de la pandemia. Como se aprecia en el Cuadro 1, los nuevos casos confirmados en España que recogen las estadísticas oficiales¹ son muy distintos en función de la fecha de publicación de las estadísticas. Así, como se aprecia en la primera columna del Cuadro 1 (la que comentan a diario los medios de comunicación), si atendemos a los Boletines que publica diariamente en PDF el Ministerio de Sanidad (excepto fines de semana donde hemos repartido a partes iguales los casos reportados los lunes), entre el 1 y el 30 de septiembre se habrían detectado en España 307.634 casos nuevos de Covid-19. Estas cifras, que reflejan los casos reportados en una fecha concreta, y que el Ministerio declara como provisionales, cambian mucho día a día una vez el Ministerio las actualiza en la web del Instituto Carlos III en función de la fecha real de realización del test, a partir de la información enviada por las CCAA a la Red Nacional de Vigilancia Epidemiológica (RENAVE). Agregando dichos datos para el conjunto de España se obtienen los casos diarios depurados y actualizados. En concreto, en la segunda columna recogemos la versión de estos datos que se publicó en el fichero [datos ccaas.csv](#) el día 8 de octubre. En este caso, entre el 1 y el 30 de septiembre la depuración y actualización de datos implicaría que en España se habrían detectado 286.781 casos nuevos de Covid-19, es decir, 20.853 casos se han perdido por el camino (un 6,8% de los reportados inicialmente).

Sin embargo, el problema no acaba aquí, ya que el 13 de octubre en la web aparece una nueva actualización de los datos (véase la tercera columna del cuadro). Uno (ingenuamente) pensaría que entre la actualización hecha el día 8 y la del 13 ya no debería haber muchos cambios en los días del mes de septiembre. Nada más lejos de la realidad, ahora los nuevos casos detectados en el mes son 299.222, es decir 12.441 más que 5 días antes. No sólo eso, si miráramos al mes de agosto (no recogido en el cuadro), comprobaríamos que entre el 8 y el 13 de octubre se han añadido también más de 4.400 nuevos casos en dicho mes.

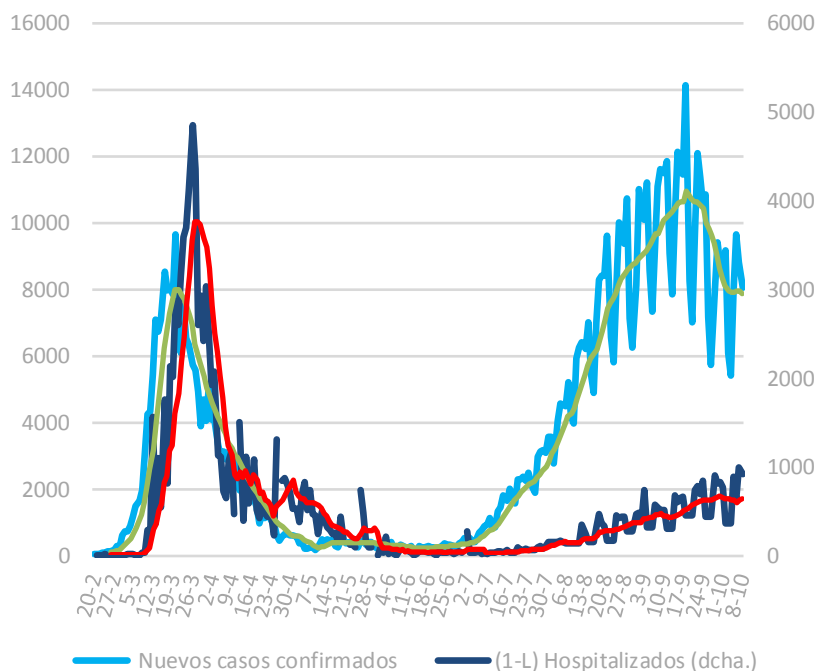
¹ Todos los datos de casos confirmados y pacientes hospitalizados que aparecen en este post provienen de estadísticas oficiales del Ministerio de Sanidad que se recogen en la siguiente página web: <https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/situacionActual.htm> Hasta la tercera semana de mayo el Ministerio publicaba datos de casos detectados, hospitalizaciones, UCI y fallecimientos en la siguiente página del Instituto Carlos III: <https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos> En la actualidad sólo publica las series completas desde el principio de la pandemia de fallecimientos y casos detectados. Las series de hospitalizados se han completado desde mayo a partir de los Boletines diarios que publica el Ministerio en la dirección web mencionada anteriormente.

Cuadro 1			
Fecha	Nuevos casos Boletín	Nuevos Casos Internet 08/10	Nuevos casos Internet 13/10
01/09/2020	8581	10720	11021
02/09/2020	8959	9916	10204
03/09/2020	10476	9751	10077
04/09/2020	8853	10870	11201
05/09/2020	8853	8241	8585
06/09/2020	8853	7061	7318
07/09/2020	8964	8822	9200
08/09/2020	8866	10717	11083
09/09/2020	10764	11219	11606
10/09/2020	12183	11123	11480
11/09/2020	9135	11416	11868
12/09/2020	9135	8721	9101
13/09/2020	9135	7514	7866
14/09/2020	9437	9851	10213
15/09/2020	11193	11730	12156
16/09/2020	11291	11400	11773
17/09/2020	14389	11046	11450
18/09/2020	10476	13752	14142
19/09/2020	10476	7856	8251
20/09/2020	10476	6712	6999
21/09/2020	10799	9258	9669
22/09/2020	11289	11717	12101
23/09/2020	10653	10999	11395
24/09/2020	12272	10124	10584
25/09/2020	10595	10396	10845
26/09/2020	10595	6768	7168
27/09/2020	10595	5212	5744
28/09/2020	9906	7232	7983
29/09/2020	11016	8666	9415
30/09/2020	9419	7971	8724
01/10/2020	11325	7468	8539
02/10/2020	7827	8163	9164
03/10/2020	7827	4858	6103
04/10/2020	7827	3725	5422
05/10/2020	11998	4978	7932
06/10/2020	10491	5084	9657
07/10/2020	12423	2311	8825
08/10/2020	12788	6	8068
09/10/2020	9285		6878
10/10/2020	9285		3040
11/10/2020	9285		1999
12/10/2020	7118		814
13/10/2020	11970		4
IA del 1 al 15 sept.	304.4	313.5	324.8
IA del 16 al 30 sept.	348.7	295.3	310.5

Obviamente diferencias como las aquí señaladas suponen cambios muy importantes en el cómputo de indicadores como el Índice Acumulado de casos por 100.000 habitantes (IA), dependiendo de si un investigador ha elegido una u otra de las tres bases de datos. Recordemos que estar por arriba o por debajo de este índice puede condicionar el confinamiento de una ciudad. En las últimas filas del Cuadro 1 se incluye esta información. Por ejemplo, entre el 16 y el 30 de septiembre el IA oscila entre 295,3 y 348,7, diferencia nada desdeñable si se estableciera un umbral por encima del cual se han de adoptar ciertas medidas. Si esto es así con los datos del conjunto de España donde las diferencias positivas y negativas pueden compensarse, ¿qué no estará pasando en las Comunidades Autónomas? ¿Van los posibles sesgos en la misma dirección en todas ellas? ¿Qué administraciones son las responsables de los posibles retrasos en la información?

Supongamos ahora que damos por buena la última de las actualizaciones disponibles, es decir, en nuestro caso la del 13 de octubre. Como es obvio mirando la última columna del Cuadro 1, debemos descartar, al menos, las cinco últimas observaciones, pues la experiencia aconseja considerar que son poco fiables ya que no recogen casos que seguro que se van a ir incorporando en un futuro en esas fechas. Por tanto, aunque vamos a trabajar con la actualización de datos del Ministerio del 13 de octubre, sólo presentaremos en los próximos gráficos datos hasta el 8 de octubre.

Gráfico 1: España. Casos detectados y variación acumulados en hospital



Fuente: Ministerio de Sanidad

Como se aprecia en el Gráfico 1, los nuevos casos confirmados que recogen las estadísticas oficiales se mantuvieron desde el final del confinamiento hasta finales de junio en niveles muy bajos (menos de 300 casos diarios en promedio en toda España). Sin embargo, desde julio y hasta la tercera semana de septiembre la curva de nuevos

contagios detectados creció hasta alcanzar números diarios que han sobrepasado con creces los de los momentos más duros del mes de marzo. Las buenas noticias son que, viendo la media móvil a 7 días (línea verde), parece que la curva de nuevos casos se está empezando a doblar, aunque éste puede ser un efecto que desaparezca si en próximas fechas se siguen reportando por parte de las Comunidades Autónomas casos imputados a esos días. En el Gráfico 1 se recoge también la evolución diaria del acumulado de pacientes hospitalizados (es decir, los nuevos ingresados, menos las salidas hospitalarias por altas o defunciones) y la media móvil (en rojo). El gráfico muestra que la serie de variación de los hospitalizados ha seguido durante la segunda ola un perfil muy distinto al de las semanas de colapso hospitalario.

¿Cuál es la razón de esta evolución tan distinta entre número de contagiados y hospitalizados entre la primera y la segunda oleada? Frecuentemente se aduce que desde el final de la primera oleada del virus se realizan más pruebas, por lo que se detectan más infectados. ¿Se puede homogeneizar en el tiempo el número de casos de contagiados efectivos teniendo en cuenta las pruebas realizadas? ¿Nos ayudaría a entender mejor la evolución de la enfermedad corregir la serie de casos detectados?

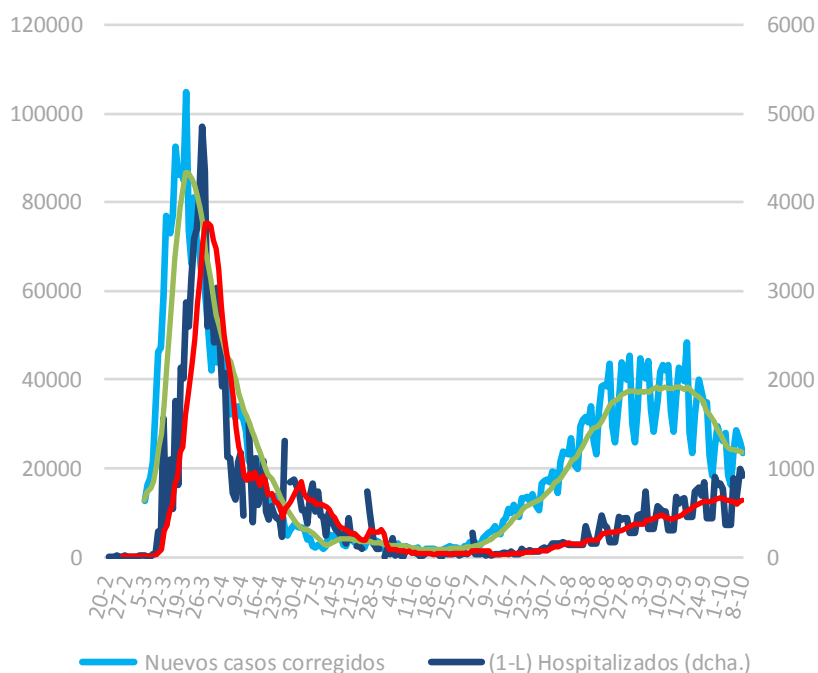
Una corrección sencilla, con ánimo simplemente ilustrativo², consistiría en suponer que el número de enfermos detectados, D , es una función $g(t)$ del número de pruebas realizadas per cápita, t , por el número de reales de contagiados, C . Es decir: $D = g(t)C$. Por tanto, la *tasa de detección*, $d = D/C$, se puede expresar como $d = g(t)$. Es razonable suponer que $g(0) = 0$, $g(1) = 1$ y que $g(t)$ es una función cóncava, pues si no hacemos pruebas no detectaremos a nadie contagiado y si hacen a todo el mundo encontraremos a todos los contagiados reales. Además, hay rendimientos decrecientes, pues empezamos haciendo las pruebas a los “mejores candidatos”: personas con síntomas o que han estado en contacto con otros contagiados. Para simplificar utilizaremos la siguiente función: $g(t) = d = t^\alpha$. ¿Cómo aproximamos α ? Contamos con dos puntos de dicha función $g(0) = 0$ y $g(1) = 1$. Para tener un tercer punto vamos a utilizar, primero, los datos de pruebas PCR realizadas a la población española que publica semanalmente desde mitad de abril el Ministerio de Sanidad y, segundo, la ratio de detectados sobre contagiados obtenido tras la primera ronda del “Estudio Nacional de Seroepidemiología de la Infección por SARS-COV-2 en España”. En dicho estudio se concluía que el 11 de mayo (último día de la primera ronda del estudio) había en España un 5% de la población contagiada, mientras que en dicha fecha se habían detectado unos 217.000 casos. En otras palabras, se habían detectado un 9,2% de los individuos realmente contagiados. Es decir, para el 11 de mayo se dispone del dato de d y t , lo que nos permite calcular un $\alpha = 0,73$.

Aplicando la anterior corrección $C = D/t^\alpha$ a la curva de casos detectados del Gráfico 1 obtendríamos la serie de contagios diarios reales estimada recogida en el Gráfico 2. La mera inspección visual de las series de contagios estimada y su comparación con la variación del acumulado de hospitalizados desvela información interesante (nótese el cambio de escala con respecto al Gráfico 1). Primero, entre el inicio de la pandemia y el 11 de mayo nuestra estimación implica que en España hubo 2.305.000 contagiados (un 4,9% de la población) en lugar de los 217.000 detectados oficialmente y que en el pico de contagios a mediados de marzo hubo cerca de 100.000 contagiados diarios. Segundo, desde el 12 de mayo hasta la actualidad en España se habrían contagiado 2.619.000 personas más y el pico diario estaría ligeramente por encima de los 40.000 casos

² Agradecemos las discusiones con Ángel de la Fuente y sus oportunas sugerencias, que han ayudado a mejorar de forma significativa esta sección.

diarios. Tercero, comparando con la primera ola, la variación de hospitalizados presenta en la segunda oleada del virus una dinámica más coherente con los casos corregidos que con los detectados. No obstante, aunque el mayor número de pruebas sí que explica una parte de esa discrepancia entre los contagiados y las hospitalizaciones, sigue existiendo una brecha importante entre la evolución de ambas magnitudes entre la primera y la segunda oleada.

Gráfico 2: España. Casos corregidos (contagiados estimados) y variación acumulados en hospital



Fuente: Ministerio de Sanidad y elaboración propia.

Un candidato obvio a explicar la diferencia es la composición por edades de los contagiados detectados en ambas olas de la pandemia. Se podría estimar una relación entre hospitalizados y contagiados en un momento anterior del tiempo controlando por la pirámide de edad. En una segunda etapa se podría utilizar dicha estimación para recalcular la serie de contagiados ya corregida por el número de pruebas, convirtiéndola en una serie de contagiados homogeneizada por edad. En este post nos limitaremos a ilustrar la muy probable relevancia de controlar por la pirámide de edad y dejando la realización del ejercicio completo para los epidemiólogos. En el Cuadro 2 se presenta información del Ministerio de Sanidad del porcentaje de casos de contagios detectados por grupos de edad, antes y después del 10 de mayo. Los datos hablan por sí mismos: la mediana de edad durante la primera ola fue de 60 años (el rango intercuartílico entre 46 y 78 años), mientras que en la segunda está siendo de 39 años (rango entre 23 y 55). Sin ninguna duda, controlar por la edad permitiría explicar una parte de la brecha que todavía existe entre casos corregidos por las pruebas y hospitalizaciones presente entre la primera y la segunda olas del virus.

Cuadro 2: Porcentaje de casos por edad.		
Grupo de edad	Entre inicio pandemia y el 10 de mayo	Entre el 11 de mayo y el 13 de octubre
<2	0,2	1,6
2-4	0,1	2,3
5-14	0,3	8,7
15-29	6,0	22,3
30-39	9,4	15,9
40-49	14,6	16,9
50-59	17,8	13,6
60-69	14,4	8,2
70-79	13,7	5,0
>80	23,5	5,5
Mediana edad	60	39
Rango	46-78	23-55

Unas últimas reflexiones para concluir. Primero, insistimos de nuevo en lo que ya se ha dicho hasta la saciedad. La calidad de los datos epidemiológicos del COVID-19 en España dista mucho de ser la deseable, sobre todo cuando se compara con otros países de la UE. El Cuadro 1 ilustra lo fácil que es tener enormes diferencias diarias e incidencias acumuladas para el mismo periodo de 50 casos arriba o abajo dependiendo de las distintas actualizaciones. Lamentablemente, gasolina para el fuego de la absurda confrontación política en la que vivimos. Segundo, se está vendiendo en los medios que hemos sobrepasado con creces la primera oleada de casos contagiados. Muy probablemente no es así. Corrigiendo por pruebas y composición de la población todavía nos sitúa en el conjunto de España lejos de dicha situación. Análisis en la dirección del presentado en este post, aunque más depurados, ayudarían a anticipar mejor hacia dónde va a transitar la pandemia en fechas venideras. Tercero, parece que ahora puede salir adelante una iniciativa del Ministerio de Sanidad para establecer un conjunto de indicadores (con niveles por colores que reflejen objetivamente la gravedad de la situación) para que todas las administraciones puedan tomar medidas anticipadamente. Bienvenida sea esta iniciativa. Nosotros mismos (y no hemos sido ni mucho menos los únicos) ya [propusimos esto en mayo](#). Reproducimos en el Cuadro 3 nuestra propuesta de entonces. Nunca es tarde si la dicha acaba siendo buena. Esperemos que sea así.

Cuadro 3

	Nuevos contagios	Fallecidos diarios	Hospitalizados	UCI	Número	Pruebas PCR	Seroprevalencia	Pro-Memoria	
	100.000 hab.	100.000 hab.	100.000 hab.	100.000 hab.	Reproductivo	1000 hab.	%	Camas Hosp.	Camas UCI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Andalucía	0.30	0.05	3.56	0.88	1.11	18.6	2.7	199	10.0
Aragón	1.74	0.16	-	-	1.13	32.9	4.9	300	10.2
Asturias	0.31	0.18	-	-	0.75	80.7	1.8	302	13.2
Baleares	0.67	0.11	-	-	1.59	61.8	2.4	251	13.6
Canarias	0.17	0.03	-	-	0.85	42.8	1.8	236	12.9
Cantabria	0.66	0.12	-	-	0.76	65.5	3.2	237	9.5
Castilla La Mancha	1.73	0.63	18.42	4.15	1.01	35.5	10.8	200	6.5
Castilla y León	2.98	0.27	-	-	0.89	55.7	7.2	290	9.8
Cataluña	1.27	0.67	-	4.14	0.65	53.3	5.9	227	8.6
Ceuta	0.51	0.00	-	-	0.87	14.9	1.1	--	--
C. Valenciana	0.63	0.10	7.57	1.26	0.93	37.5	2.5	204	9.0
Extremadura	0.45	0.23	-	-	0.80	30.3	3	265	7.0
Galicia	0.29	0.05	4.02	0.71	0.67	54	2.1	290	6.1
Madrid	2.45	0.37	28.51	5.36	0.74	69.4	11.3	228	9.7
Melilla	0.33	0.00	-	-	2.18	20.4	1.9	--	--
Murcia	0.39	0.06	-	-	0.99	21.9	1.4	235	8.9
Navarra	2.64	0.22	-	-	0.90	73.2	5.8	247	6.5
P. Vasco	1.51	0.27	15.23	2.44	0.74	81.2	4	240	8.2
La Rioja	0.86	0.23	-	-	1.13	93.6	3.3	196	5.4
España	1.14	0.26	-	-	0.81	47.2	5	232	9.3