

EL NOMBRE DE REPRODUCCIÓ DE LA COVID-19 I EL MODEL SIR. L'EFECTE DELS RETARDS DE COMPTABILITZACIÓ

Xavier Mora

Dep. Matemàtiques, UAB

Esborrany 0.7, 20 maig 2020

Aquests dies se sent a parlar molt del nombre de reproducció R d'una epidèmia. Tot sovint es parla de R_0 en lloc de R , però en les condicions que ens interessin és essencialment el mateix. El que importa és que per a aturar una epidèmia cal mantenir R per sota de 1, la qual cosa vol dir unes mesures adequades per a evitar suficientment els contagis; altrament, l'epidèmia creix sense aturador. Davant d'això és essencial mesurar aquest paràmetre a partir de les dades existents. Tanmateix, fàcilment ens trobem amb discrepàncies importants. Per exemple, segons un informe del Ministerio de Sanidad [16] el dia 1 de maig a Catalunya teníem un valor de 0,69. En canvi, segons *El País* [8], el valor a Catalunya en la mateixa data era aproximadament 1. I uns altres càlculs que presentem més avall donarien un valor de l'ordre de 2. Com pot ser que hi hagi aquestes divergències? Quin és el valor correcte?

1 Definicions i primeres reflexions

Comencem per la definició. *El nombre de reproducció R és el nombre mitjà de casos nous a què dona lloc per contagi directe cada cas existent.* Noti's que no es diu res del moment en què apareguin aquests casos nous. No es tracta, doncs, d'una 'taxa' temporal —casos per unitat de temps— sinó que compten per igual tots els casos que tenen un mateix precursor immediat, independentment de quan es produeixen. És per això que R no té unitats i es parla d'aquest paràmetre com a 'nombre', o 'ratio', o 'factor'; però no 'taxa'. En lloc de 'nombre de reproducció' en podem dir també 'factor de multiplicació'. D'altra banda, tot i que no sigui una taxa temporal, el valor de R pot variar amb el temps, a conseqüència de variacions en les condicions que faciliten o dificulten els contagis.

Pel que fa a R_0 , que es coneix com a nombre bàsic de reproducció, en principi es refereix al cas particular d'una possible epidèmia que està començant amb un nombre d'infectats molt petit en comparació amb tota la població (la qual cosa valdria també ara, que la població infectada no arriba a l'1%). D'altra banda, la denominació R_0 també acostuma a suposar unes

condicions ‘naturals’ en relació amb el contagi, que no és el cas quan s’han adoptat mesures per a prevenir-lo.

D’on poden provenir, doncs, les discrepàncies en el valor de R ? A continuació considerem diferents possibilitats, no necessàriament independents entre elles.

(1) La primera és què entenem per un *cas* de la malaltia. Considerem només els casos greus? O potser només els que han estat confirmats per una determinada prova de diagnòstic? Includim també els sospitosos? Potser hem de considerar també els asimptomàtics? Potser no és tan estrany que, segons quin criteri adoptem en aquest punt, obtinguem valors diferents de R .

(2) Davant la possibilitat d’infecciosos asimptomàtics, o amb símptomes lleus, certament és natural fer campanyes generalitzades de proves de diagnòstic. Tanmateix, això complica les coses, ja que llavors cal distingir entre els casos que apareixen com a “nous” a causa d’estar fent proves, i els que són realment “nous” per la dinàmica de la malaltia. Si confonem les dues coses, llavors podem obtenir uns valors de R més alts del compte.

(3) D’altra banda, tot i que R no sigui una taxa temporal, les dades sí que estan distribuïdes en el temps. Per tant, a l’hora d’estimar el valor de R a partir de les dades és poc menys que inevitable fer alguna hipòtesi sobre el mecanisme de propagació de la malaltia al llarg del temps, que pot ser molt variat. Vegi’s, per exemple, [5]. Això també pot ser una font de discrepàncies. Dins d’aquest apartat també hi podem incloure el punt de vista de [3, 7], que prescindeix del mecanisme però requereix una hipòtesi sobre el temps que media entre un cas i els que deriven directament d’ell.

2 Sobre les dades

En el cas de Catalunya, tenim les dades que publica cada dia al vespre el Departament de Salut de la Generalitat [12] i les que publica al matí del dia següent el Ministerio de Sanidad [14]. Tot i que se suposa que les segones provenen de les primeres, els valors són diferents, a causa d’una diferent definició de ‘cas’. Com es veurà més avall (figures 2, 3 i 11) això resulta en diferències importants en l’estimació del valor de R ; en particular, de vegades passa que unes dades donen $R < 1$ i les altres $R > 1$.

De tant en tant, les dades presenten unes irregularitats que cal tenir en compte. Com ara que la xifra acumulada d'un dia sigui inferior a la del dia anterior. Això ha passat un parell de vegades degut a un canvi de criteri en la definició de ‘cas’. En tals situacions, l’Institut de Salut Carlos III (ISCI) [14] —que és l’entitat que serveix les dades al Ministerio de Sanidad— procura revisar tot l’historial d’acord amb el nou criteri, la qual cosa facilita l’anàlisi.

De vegades també s’hi revisa la data dels esdeveniments —casos, altes, morts— que han estat notificats amb retard. En particular, la Comunidad de Madrid revisa sistemàticament les dades de dies anteriors a fi de col·locar cada cas confirmat en la data en què es va prendre la mostra o es va obtenir el resultat de la prova.

Malgrat aquestes revisions, es mantenen certs patrons setmanals que evidencien uns retards sistemàtics associats amb els caps de setmana. Vegi’s, per exemple, la figura 1 de [13], la qual reproduïm a continuació. Davant d’això es fa molt convenient sotmetre les dades a una primera operació de mitjanes mòbils de 7 dies, la qual cosa és força habitual a la literatura: el valor del dia t el substituïm per la mitjana dels valors dels dies $t - 3, \dots, t + 3$; això per a cada t i cada tipus de dades (casos, altes i morts). Noti’s, però, que a conseqüència de fer aquestes mitjanes, els resultats més recents que podem donar no són d’avui, sinó de fa 4 dies.

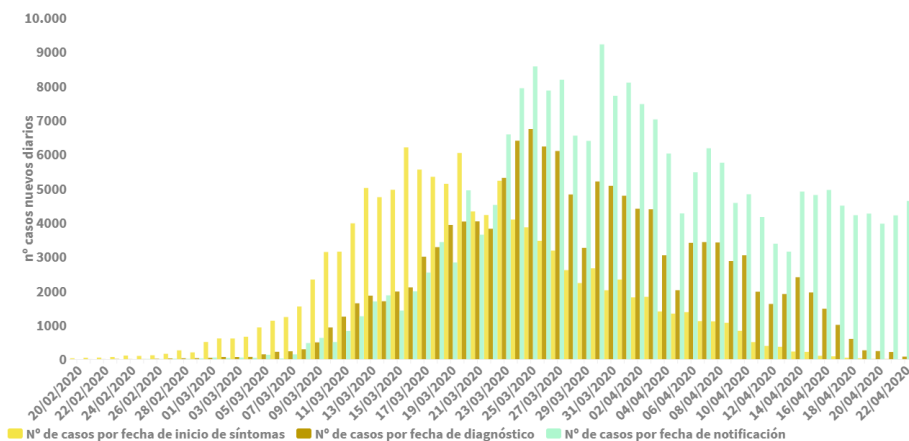


Figura 1: “Casos diarios confirmados de COVID-19 por fecha de notificación, diagnóstico e inicio de síntomas. 23.04.2020 (datos consolidados a las 21:00 horas del 22.04.2020)” [13].

3 Estimació basada en el model SIR

En principi, l'evolució del nombre de casos de COVID-19 s'hauria de poder modelar, si més no aproximadament, mitjançant les equacions diferencials que van introduir William O. Kermack i Anderson G. MacKendrick el 1927 i que actualment es coneixen com a model SIR [1, 4].

Aquest model considera les següents variables d'estat (per evitar que R tingui dos significats diferents, la tercera variable l'anomeno X en lloc de R , tal com es fa també a [2]):

S nombre de persones susceptibles de contraure la malaltia,

I nombre de persones que tenen la malaltia (encara que no es manifesti) i la poden transmetre,

X nombre de persones que han passat la malaltia (altes o morts) i ja no la poden transmetre.

Aquestes variables varien amb el temps d'acord amb les següents equacions diferencials, on el producte $\beta S I$ modela la velocitat amb què apareixen nous infectats, i el producte γI modela la velocitat amb que els infectats deixen de ser-ho:

$$dS/dt = -\beta S I, \quad (1)$$

$$dI/dt = \beta S I - \gamma I = (\beta S - \gamma) I, \quad (2)$$

$$dX/dt = \gamma I. \quad (3)$$

Noti's que $d(S + I + X)/dt = 0$, de manera que el total $N := S + I + X$ es manté constant.

El nombre de reproducció R està relacionat amb els paràmetres β i γ de la manera següent (vegi's, per exemple, [1, p.353]):

$$R = \beta S / \gamma. \quad (4)$$

Noti's que la desigualtat $R > 1$ és exactament equivalent a $\beta S - \gamma > 0$, i per tant a $dI/dt > 0$. (Per cert, $\beta S - \gamma$ és la taxa temporal de creixement del nombre de casos actius).

Segons es dedueix de les equacions (1)–(3), els paràmetres β i γ es poden estimar a partir de les dades de la manera següent:

$$\beta S = \frac{d(I + X)/dt}{I}, \quad (5)$$

$$\gamma = \frac{dX/dt}{I}, \quad (6)$$

i per tant,

$$R = 1 + \frac{dI}{dX}. \quad (7)$$

Evidentment, a la pràctica les derivades seran substituïdes per quocients d'increments petits. Més concretament, és adient natural considerar increments diaris de les mitjanes setmanals que hem dit més amunt.

A continuació es mostren els resultats sobre R que s'obtenen per a Catalunya, segons que s'utilitzin les dades de Cat/Salut o bé les de l'ISCIH. Tal com s'ha dit més amunt, corresponen a dues definicions diferents de 'cas' i el dia 1 de maig els seus valors eren respectivament major i menor que 1. *Afegit el 19 de maig de 2020*: Els cinc valors superiors a 1 del final de la fig. 3 són un artefacte causat per retards de notificació (vegi's la nota de la pàg. 19).

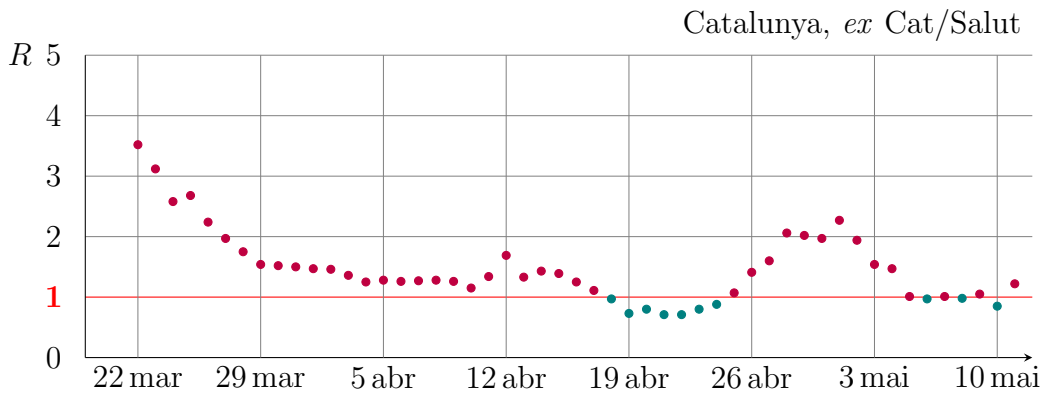


Figura 2

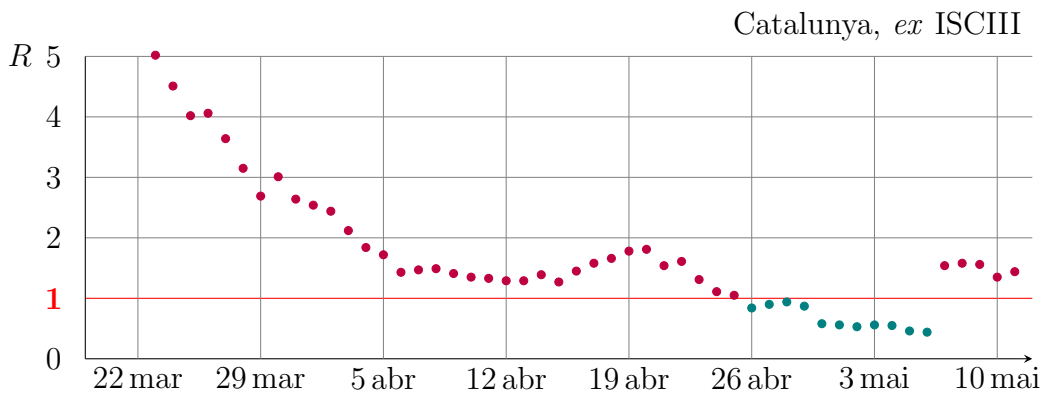


Figura 3

4 El retard entre els fets i la seva comptabilització

En fer aquestes estimacions a partir de dades reals, hom es troba amb variacions dels paràmetres β i γ al llarg del temps. Per exemple, les figures que segueixen mostren els valors de $1/(\beta S)$ i $1/\gamma$ (en dies) per a Catalunya segons les dades de Cat/Salut:

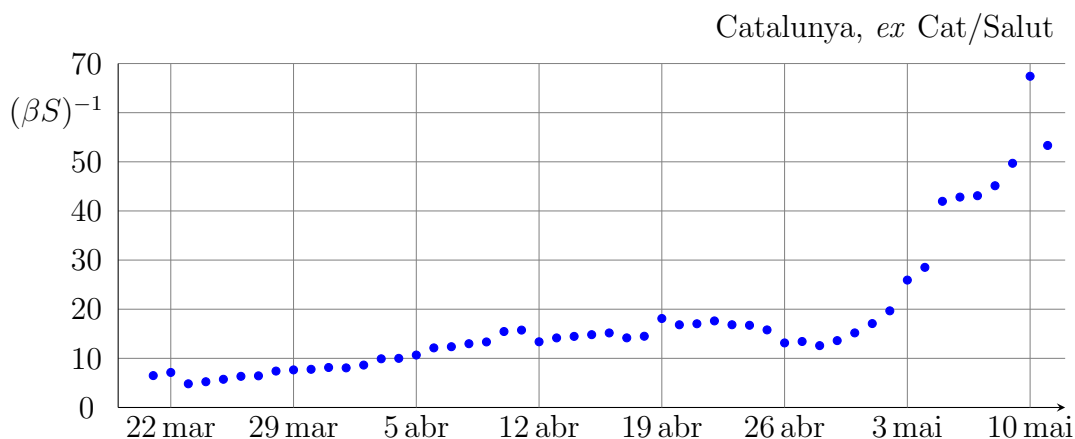


Figura 4

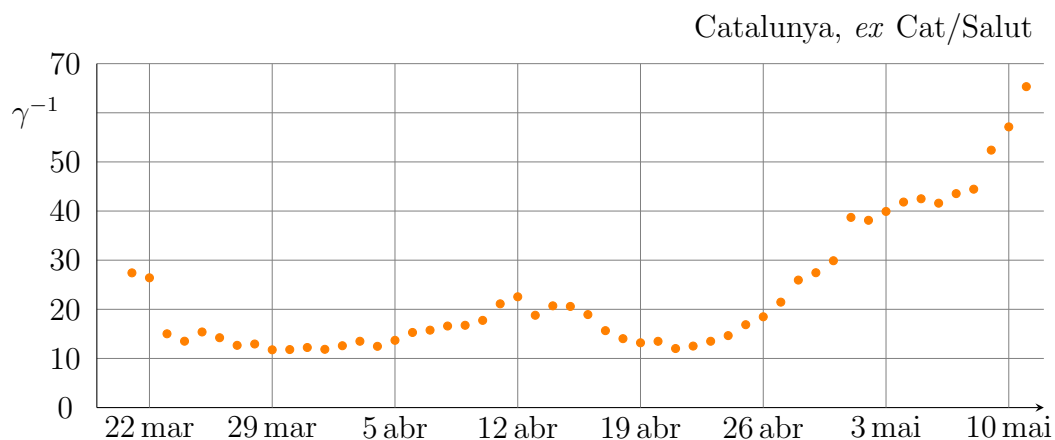


Figura 5

En el cas de β aquestes variacions poden ser atribuïdes a les variacions en la freqüència de contactes i en les mesures de prevenció. Però el paràmetre γ no és epidemiològic sinó clínic. Més concretament, $1/\gamma$ és el temps mitjà de

durada de la malaltia (i per tant, el temps mitjà que un malalt és infecciós). Tal vegada el fet que aquest paràmetre variï amb el temps té a veure amb l'eficiència del sistema sanitari. Però també cal tenir en compte una altra possible font de variació, la qual afecta tant γ com β .

En efecte, d'acord amb el que hem vist a la secció 2, podem distingir dues variables de temps:

t temps en què tenen lloc els fets (nous casos, altes, morts),

s temps en què aquests fets són comptabilitzats.

Òbviament, $s \geq t$. La diferència $s - t$ és el retard entre el fet i la seva comptabilització.

La figura 1 posa de manifest que aquest retard és de l'ordre de quinze dies o més! Aquesta afirmació es basa en comparar el màxim del nombre de casos nous per data de notificació (barres verdes) amb el màxim del nombre de casos nous per data d'inici de símptomes (barres grogues). Com que les dades no són completes, no podem estar segurs que la barra verda més alta correspongui a la barra groga més alta; potser la més alta d'un color correspon a la segona més alta de l'altre. De tota manera, queda prou clar que hi ha un retard de l'ordre de 10 o 14 dies entre l'inici de símptomes i la comptabilització del cas. A això cal sumar-hi encara uns 5 dies més d'incubació després del contagi. En total, resulta que els casos que van ser reportats com a “nous” el dia 30 de març, en realitat havien aparegut uns 10–14 dies abans, i els contagis s'havien produït encara uns 5 dies més enrera!

Certament, abans que un cas quedi comptabilitzat ha calgut concertar una cita amb el servei mèdic, receptor una prova de diagnòstic, recollir-ne el resultat, notificar-lo a on correspongui i transmetre la informació als centres de control autonòmics i estatal. Per tant, aquest retard pot ser més o menys comprensible. Tanmateix, cal tenir-lo molt en compte a l'hora d'interpretar les dades que ens arriben cada dia sobre el nombre de casos “nous”. En particular,

Cal tenir en compte aquest retard d'uns quinze dies a l'hora de comprovar els efectes de les mesures de confinament o desconfinament.

D'altra banda, el retard $s - t$ entre fets i comptabilització no té perquè ser una constant. I les seves possibles variacions en el temps influeixen en els valors “aparents” de β i γ . Perquè la dinàmica de l'epidèmia, i el model SIR, fan referència a t . Però les dades que es publiquen fan referència a s .

I una derivada com dX/ds , que és el que observem en el recompte diari de noves altes i morts, no és γI , sinó $(dt/ds)\gamma I$. Per tant, el valor que obtenim mitjançant el quocient $(dX/ds)/I(s)$ —anàleg al segon membre de (6)— no és γ , sinó $(dt/ds)(I(t)/I(s))\gamma$, que en podem dir γ aparent. Així, encara que γ fos constant, si el retard és variable llavors el valor aparent de γ és variable.

En relació amb aquesta qüestió es podria pretendre reconstruir la dinàmica real de la malaltia, en la variable t . Però les dades les tenim en la variable s , i la validació del model s’ha de basar en aquestes dades. Per tant, sembla prou adient quedar-se en la variable s i admetre les variacions que llavors s’observen en el paràmetre γ . En el benentès que no estem modelant només la dinàmica de l’epidèmia, sinó que també incluïm el retard entre els fets i la seva comptabilització.

Si el recompte d’altes i de morts té el mateix retard que el recompte de casos nous, llavors els dos termes de la dreta de l’equació (2) són multiplicats per un mateix factor dt/ds , i per tant, el seu quocient no es veu afectat, de manera que roman igual a R . Però en cas contrari el quocient dels valors aparents de βS i γ es pot desviar del veritable valor de R . Més endavant (secció 8) veurem que aquestes desviacions es produeixen també encara que els retards siguin constants (però diferents) i que això explica certes observacions.

5 Altres exemples

Tal com s’ha dit més amunt, la Comunidad de Madrid revisa cada dia tot l’historial per tal de reassignar cada cas confirmat a la data en què es va realitzar la prova. Això fa que les dades quedin més endreçades i que la seva evolució temporal sigui més suau. De tota manera, el retard entre fets i comptabilització no queda eliminat del tot, sinó que resta el retard entre el contagi i la confirmació per la prova de diagnòstic, i aquest retard encara pot ser de l’ordre d’una setmana i mitja.

L’evolució de β, γ i R per a la Comunidad de Madrid es mostra a la figura 6. Les figures 7 i 8 mostren el mateix per a Espanya en conjunt i per a Singapur (dades provinents de [17, 19]). Com es pot veure, en aquest últim cas s’obtenen uns valors (aparents) de R propers a 30, la qual cosa podria ser deguda a la possibilitat esmentada més amunt de retards diferents segons que es tracti de casos nous, altes o morts.

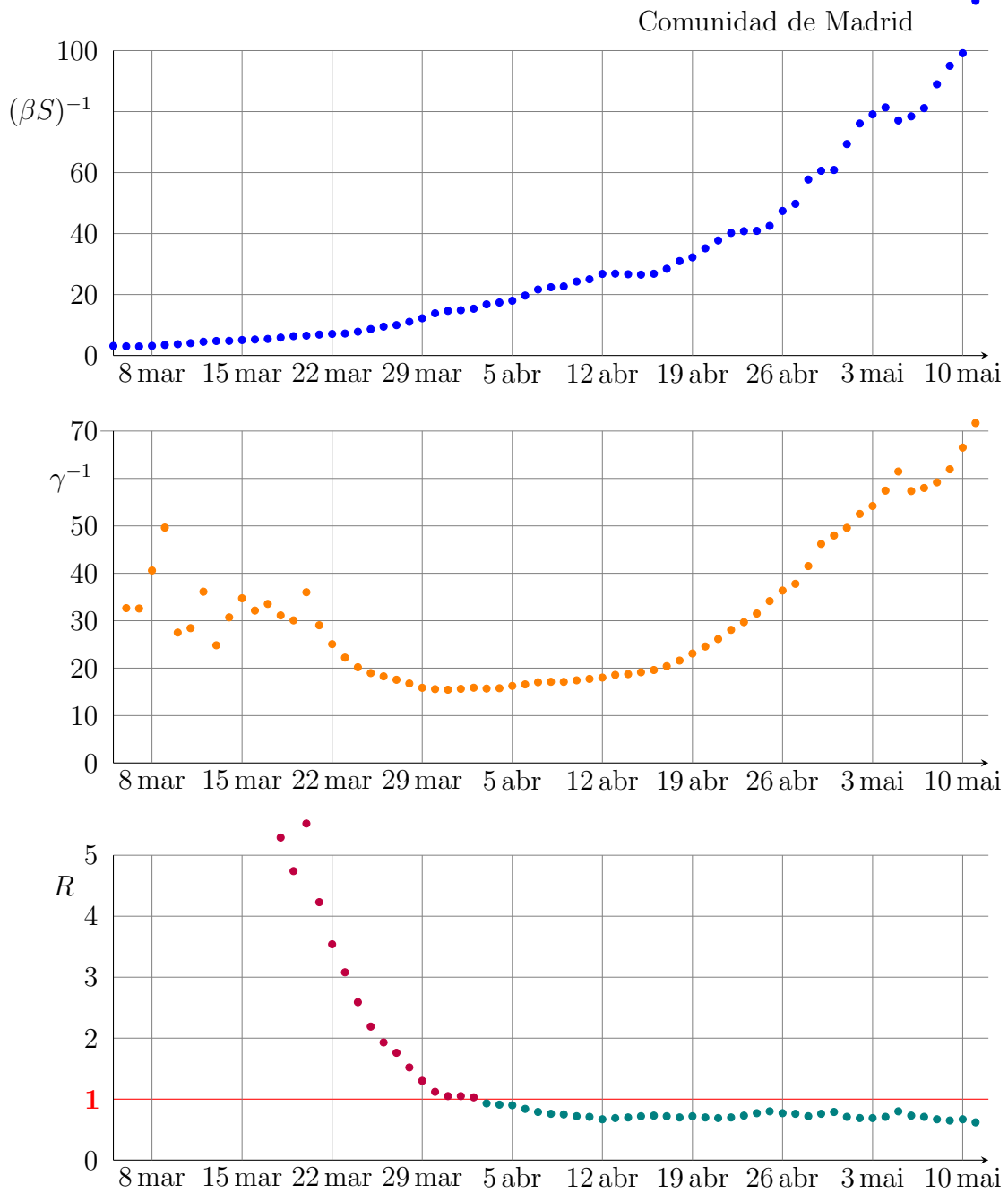


Figura 6

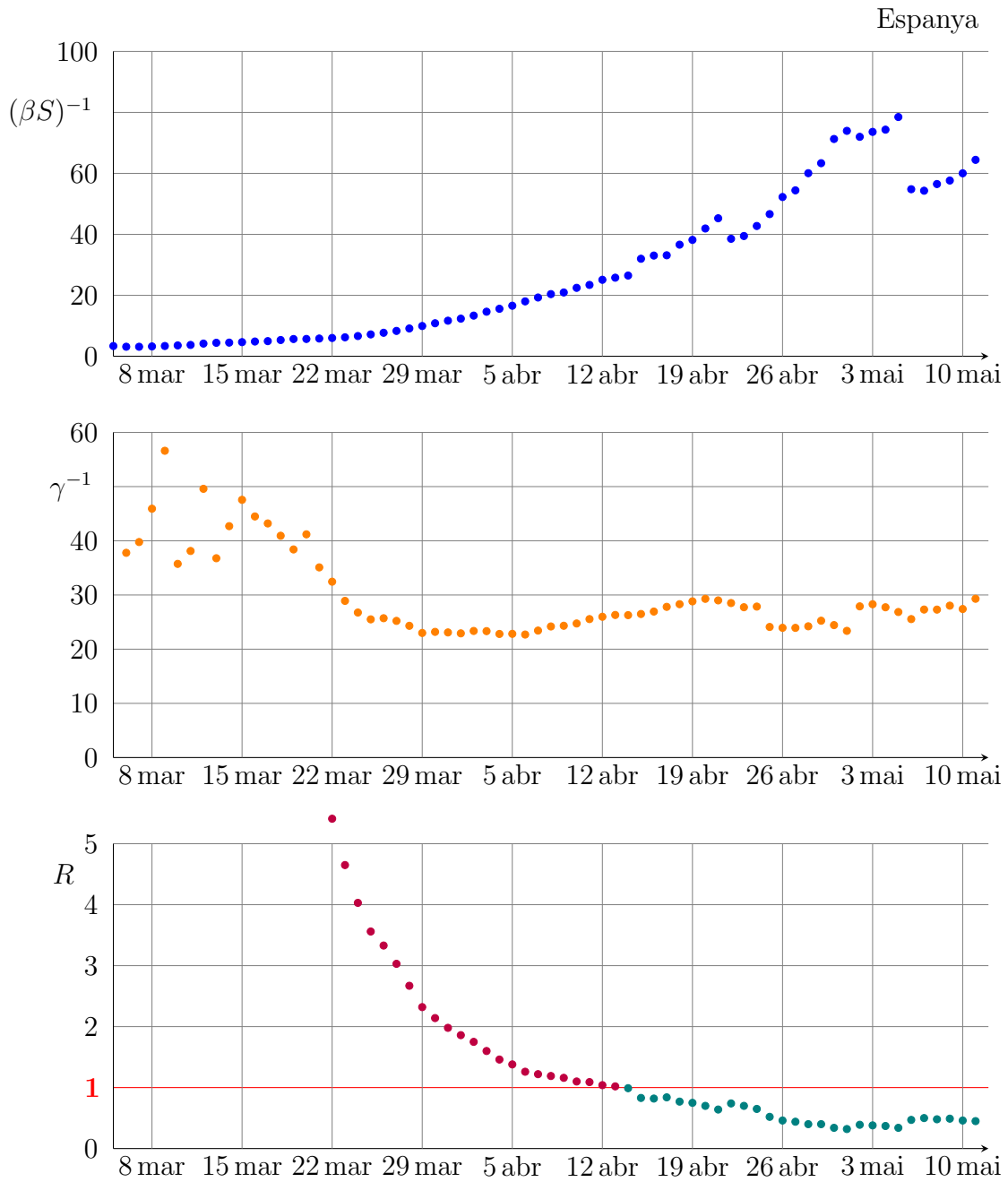


Figura 7

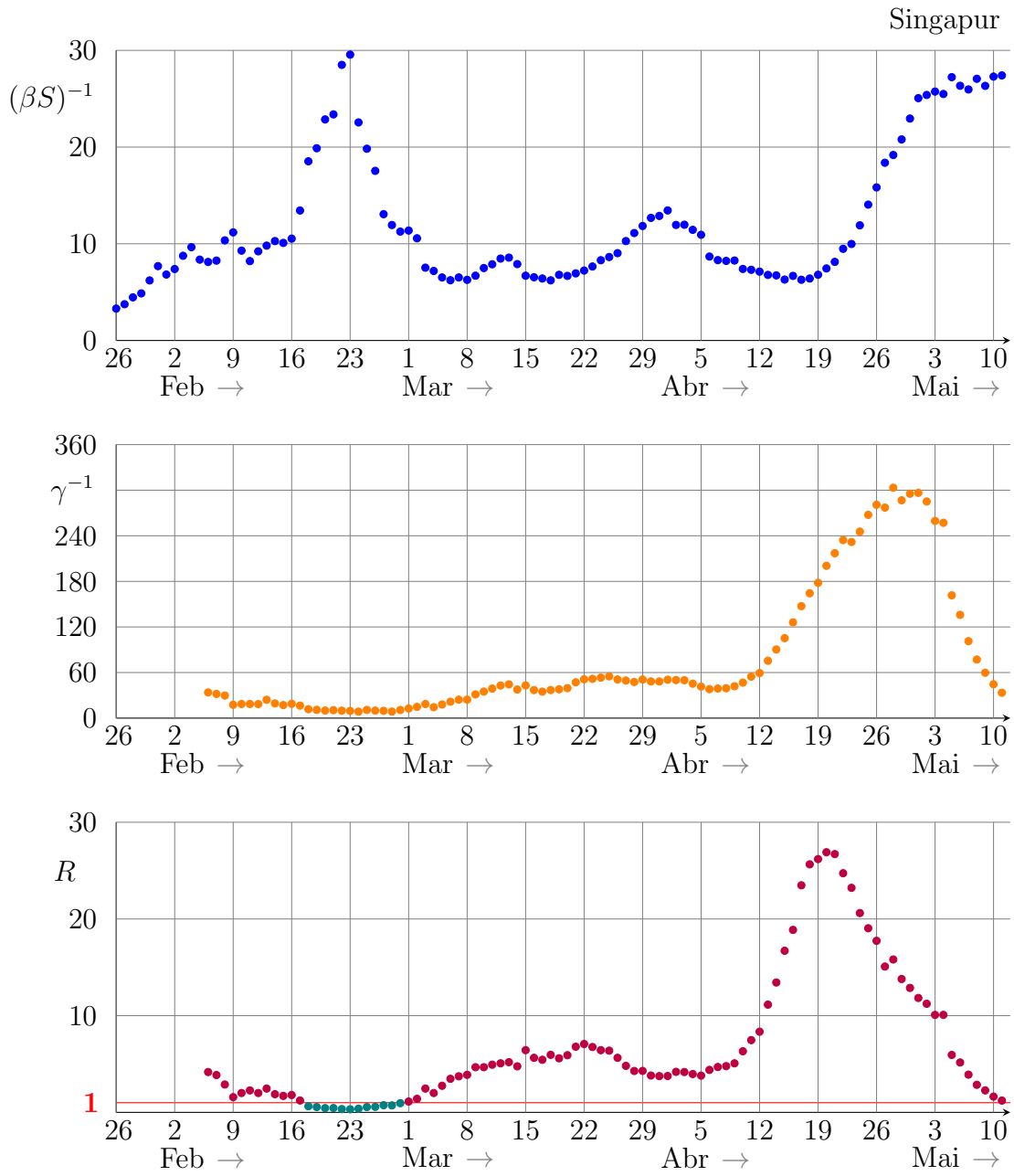


Figura 8

6 Previsions a curt termini

Si coneguéssim bé el valor dels paràmetres, el model SIR ens permetria predir l'evolució de l'epidèmia. Tanmateix, ja hem vist que els paràmetres tenen incerteses importants. Fins i tot en absència de retards, serà normal que β variï a conseqüència de les mesures de prevenció i confinament o desconfinament. Per tant, només es pot pretendre fer previsions a curt termini. A continuació mostrem aquestes previsions en els diversos casos que hem estat examinant, suposant que els paràmetres mantinguessin els seus darrers valors. En cada cas mostrem només l'evolució temporal del nombre de casos actius $I(t)$.

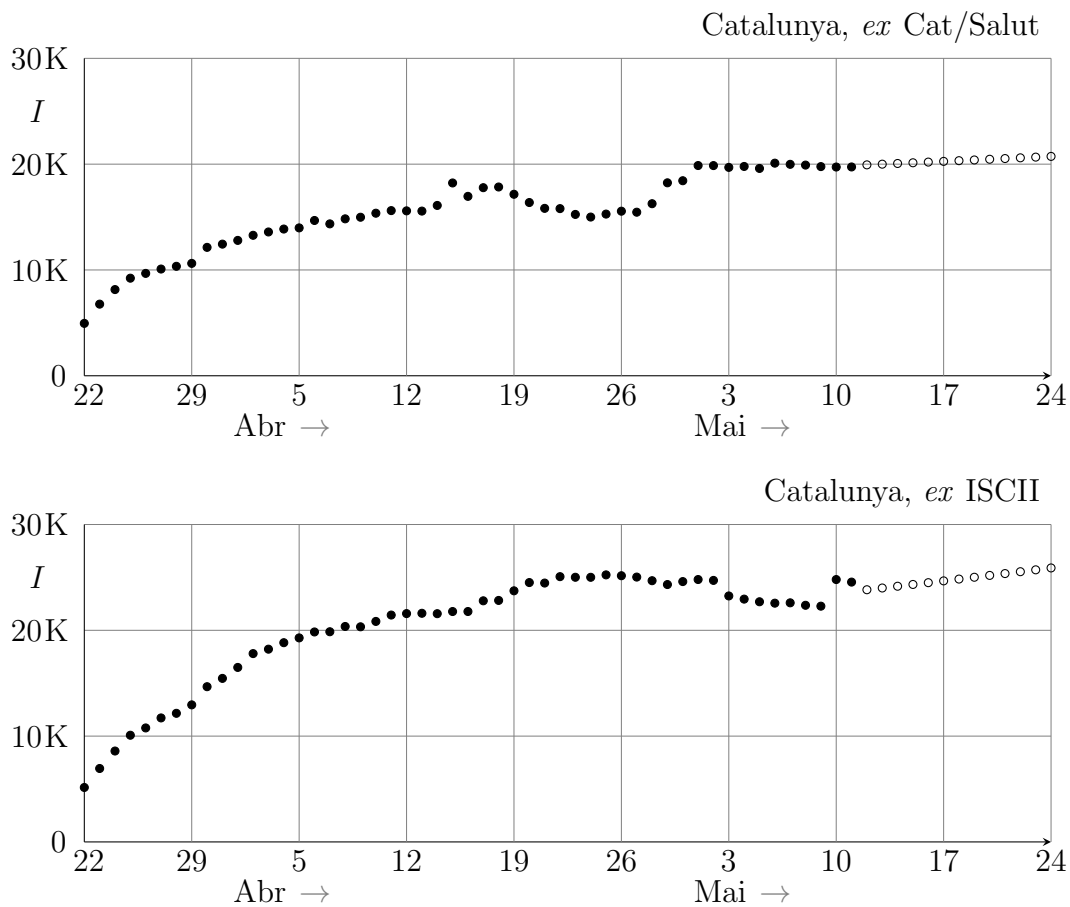


Figura 9

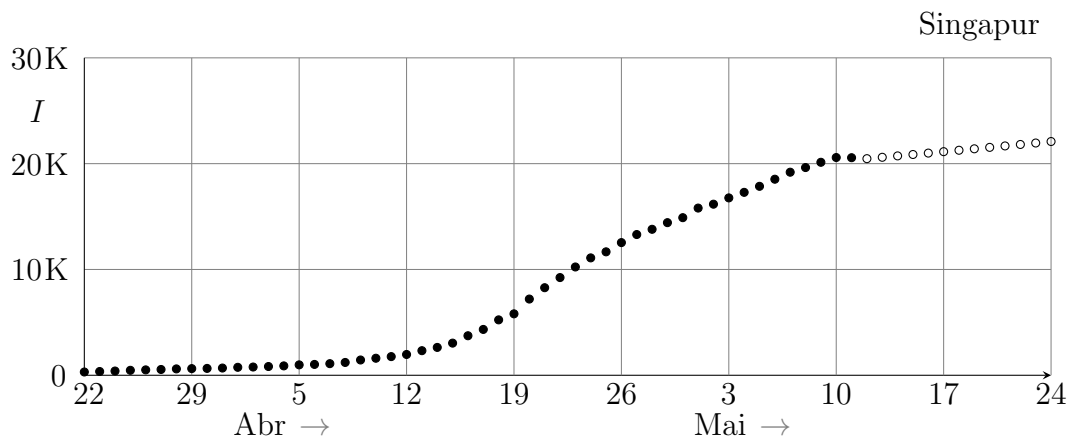
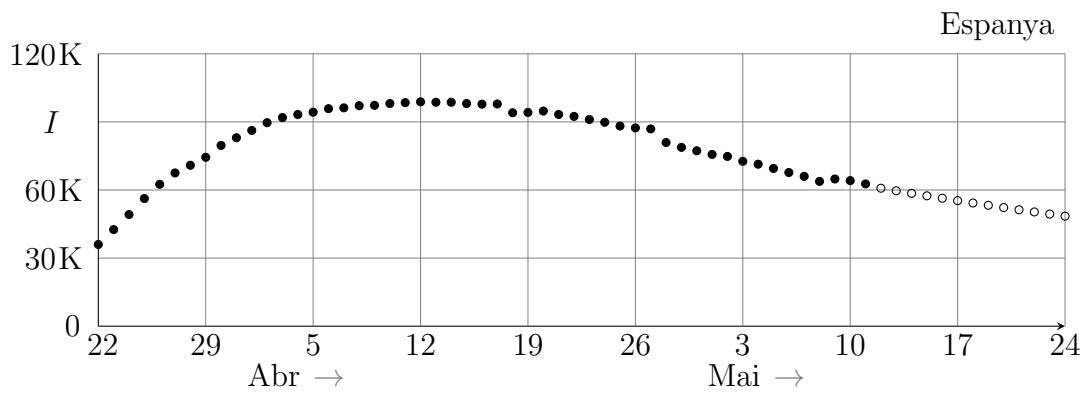
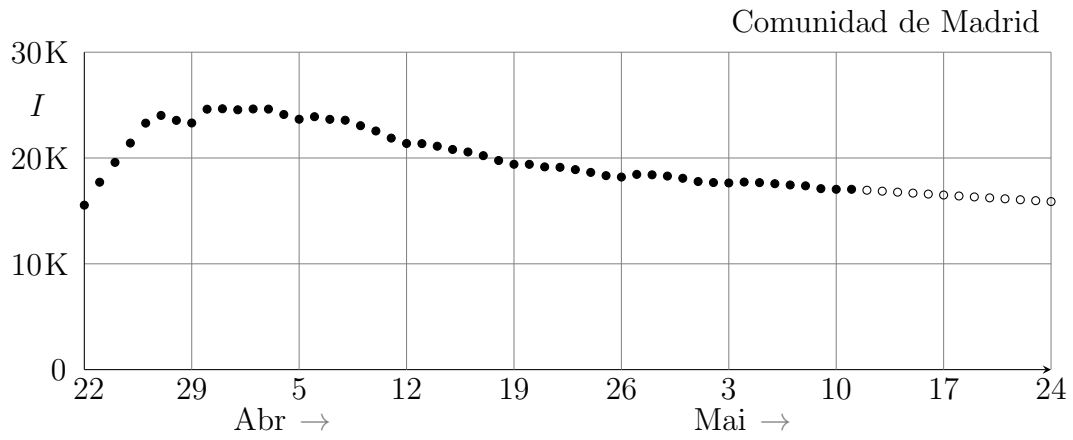


Figura 10

7 Comparació amb l'estimació de R basada en el temps de generació

La major part de les estimacions que s'estan donant actualment del paràmetre R per a la COVID-19 no es basen en un model determinista de la dinàmica, com ara el model SIR, sinó en un model probabilístic de l'anomenat temps o interval de generació.

El temps de generació vol dir el temps transcorregut entre la infecció d'una persona i la d'una segona persona contagiada per la primera. Com que el començament de la infecció sol ser difícil de situar en el temps, habitualment se suposa que aquest temps és el mateix que media entre l'aparició de símptomes de les dues persones (que pròpiament parlant es coneix com a *serial interval*). Per suposat, aquest temps pot variar d'un cas a un altre, de manera que correspon descriure la seva variabilitat mitjançant les eines habituals de la teoria de probabilitats, començant pel seu valor mitjà i la seva desviació típica.

Doncs bé, tal com es mostra a [7], resulta que aquesta informació, que pot ser obtinguda mitjançant una investigació estadística de casos concrets, permet determinar R a partir de la taxa temporal de creixement de l'epidèmia, el valor de la qual es dedueix de les dades publicades. Aquesta connexió es troba a la base de l'eina `EpiEstim` de [3], la qual és utilitzada per la major part dels estudis de seguiment de la COVID-19 [8, 9, 14]. L'estudi [10] usa una versió simplificada de la mateixa idea.

Les gràfiques que segueixen mostren els resultats d'aquest procediment d'estimació de R en comparació amb els que hem obtingut més amunt. Més concretament, hem utilitzat l'eina `EpiEstim` i hem suposat que la mitjana i la desviació típica del temps de generació valen respectivament 3.96 i 0.21 dies, els quals valors provenen d'un dels estudis que estan recollits a [6]. (Aquest valor és inferior o similar al període d'incubació, d'uns 4–6 dies segons la mateixa referència, la qual cosa concorda amb la relativa freqüència dels contagis per asimptomàtics).

Com es pot veure, les nostres estimacions basades en el model SIR acostumen a ser superiors a les que es basen en el temps de generació, i algunes vegades passa que el primer valor és superior a 1 i el segon és inferior a 1. En el cas de la Comunidad de Madrid hi ha concordància prou bona durant l'últim mes, però abans d'això les dues estimacions difereixen en un factor de l'ordre de 3. Tant en el cas de Catalunya *ex Cat/Salut* com en el de Catalunya *ex ISCIII*, s'observen diferències del mateix ordre. Tot i això, les dues trajectòries tenen una certa similitud qualitativa.

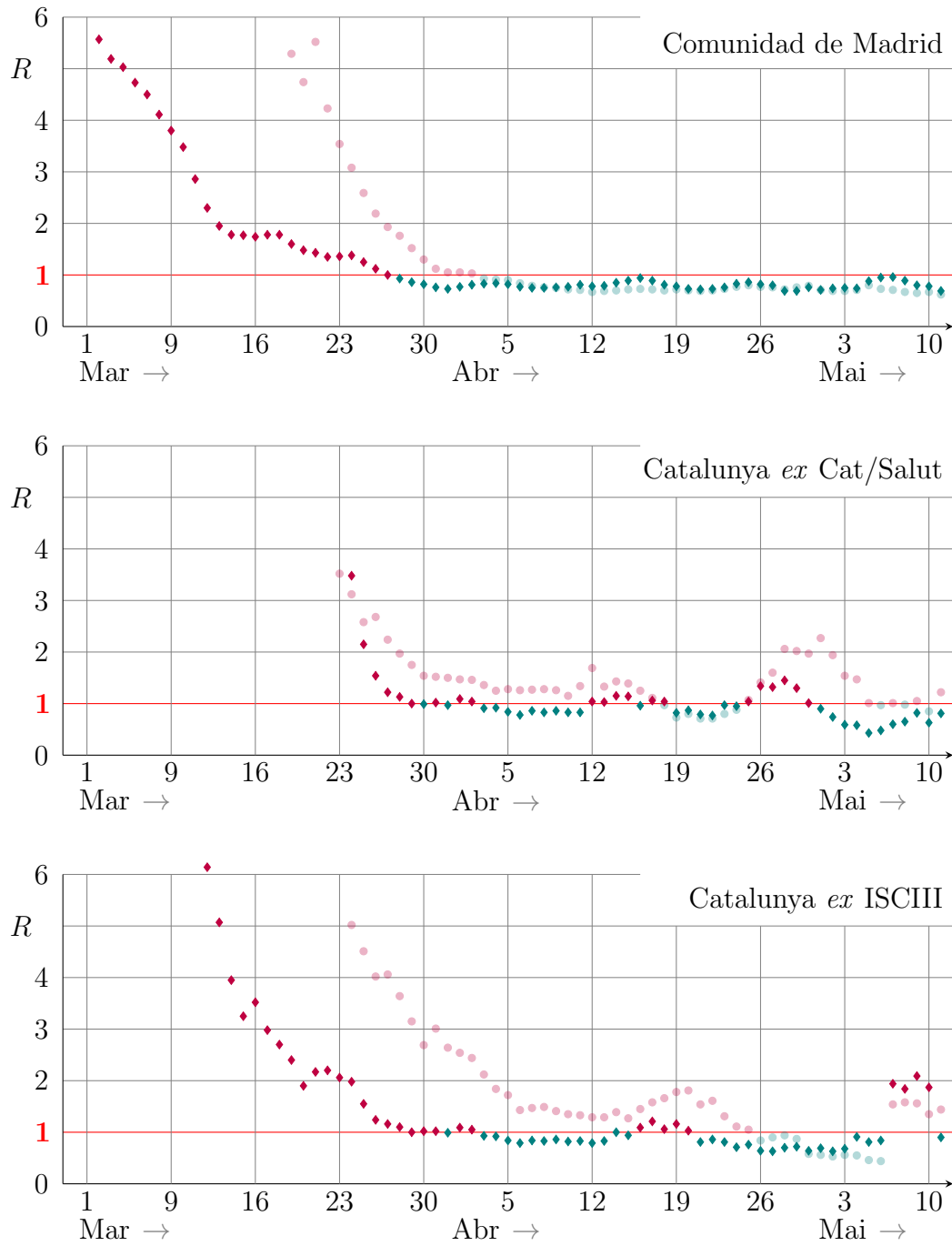


Figura 11: \blacklozenge estimació “directa”; \bullet estimació SIR

Tanmateix, en el cas de Singapur (figura 12) la ratio d'un valor a l'altre arriba a ser de l'ordre de 30! Com veurem tot seguit, tot i ser molt grans, aquestes divergències serien degudes als retards de comptabilització.

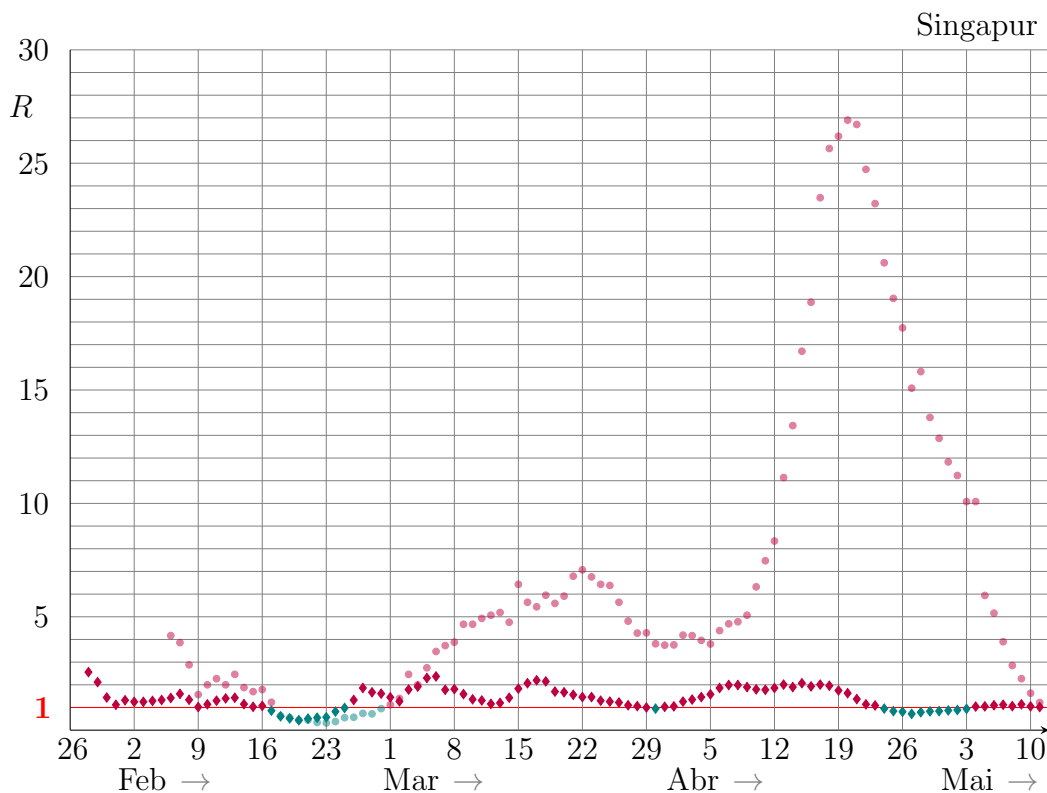


Figura 12: ♦ estimació “directa”; ● estimació SIR

8 Més sobre els retards de comptabilització

A la secció 4 hem vist com les estimacions de β i γ poden quedar distorsionades a causa d'un retard variable en la comptabilització dels fets (nous casos, altes i morts). I també hem comentat que si el retard és el mateix per als tres tipus d'esdeveniments, llavors l'estimació de R no es veu afectada.

A continuació explorem el cas en què hi hagi un retard diferent segons el tipus d'esdeveniment, la qual cosa és molt possible que passi en la realitat. Com veurem, llavors apareixen distorsions fins i tot en el cas que cadascun d'aquests retards es mantingui constant en el temps.

Per veure el que passa serà suficient considerar el següent escenari: β i γ són constants (però desconegudes); I i X són molt inferiors a S , i per tant, S és aproximadament constant; a temps 0 la variable I té un valor conegut I_0 . La comptabilització de casos nous té un retard constant p , i la comptabilització d'altres i morts té un retard constant q . Per simplificar el llenguatge no distingirem entre altres i morts.

Usarem també la següent notació: $b = \beta S$, $\delta = b - \gamma$. Per últim, els valors que es tenen en presència de retards els representarem mitjançant una notació de la forma \tilde{A} , on A representa el valor que es tindria en absència de retards.

El punt de partida són els nombres diaris de casos nous i altres noves, els quals denotarem respectivament Δ i Γ . Amb les hipòtesis que estem fent, les equacions (2)–(3) es resolen sense cap dificultat i determinen que

$$\Delta(t) = d(I+X)/dt = b I_0 \exp(\delta t), \quad (8)$$

$$\Gamma(t) = dX/dt = \gamma I_0 \exp(\delta t), \quad (9)$$

on les derivades seran a la pràctica increments diaris. Tanmateix, en lloc d'aquests valors es comptabilitzen els corresponents valors retardats, a saber

$$\tilde{\Delta}(s) = \Delta(s-p) = b I_0 \exp(\delta(s-p)), \quad (10)$$

$$\tilde{\Gamma}(s) = \Gamma(s-q) = \gamma I_0 \exp(\delta(s-q)), \quad (11)$$

on, per consistència amb la secció 4, utilitzem la notació s per a recordar que es tracta del temps de comptabilització. El procediment que utilitzem per a estimar b i γ consisteix en dividir aquestes quantitats pel nombre de casos actius —equacions (5) i (6)—. Si no hi hagués retards, el nombre que comptaríem de casos actius seria

$$I(t) = I_0 + \int_0^t (\Delta(\tau) - \Gamma(\tau)) d\tau = I_0 \exp(\delta t), \quad (12)$$

que coincideix, certament, amb la solució exacta de les equacions (2)–(3). Per suposat, a la pràctica la integral serà una suma. Però en presència de retards, i suposant, de moment, que la taxa de creixement $\delta = b - \gamma$ no és nul·la, el que comptem és

$$\tilde{I}(s) = I_0 + \int_0^s (\tilde{\Delta}(\tau) - \tilde{\Gamma}(\tau)) d\tau = \dots = I_0 (1 - Q + Q \exp(\delta s)), \quad (13)$$

on

$$Q = \frac{b e^{-\delta p} - \gamma e^{-\delta q}}{b - \gamma}. \quad (14)$$

I les estimacions que obtenim de b, γ són:

$$\tilde{b}(s) = \frac{\tilde{\Delta}(s)}{\tilde{I}(s)} = \frac{e^{\delta(s-p)}}{1 - Q + Q e^{\delta s}} b, \quad (15)$$

$$\tilde{\gamma}(s) = \frac{\tilde{\Gamma}(s)}{\tilde{I}(s)} = \frac{e^{\delta(s-q)}}{1 - Q + Q e^{\delta s}} \gamma, \quad (16)$$

d'on en resulta que

$$\tilde{R} = e^{\delta(q-p)} R. \quad (17)$$

En el cas especial $\delta = b - \gamma = 0$, l'equació (13) queda substituïda per $\tilde{I}(s) = I_0$, la qual cosa resulta en $\tilde{b} = b$, $\tilde{\gamma} = \gamma$ i $\tilde{R} = R$.

Com es pot veure, per a $\delta \neq 0$, $\tilde{b}(s)$ i $\tilde{\gamma}(s)$ depenen de s llevat que $p = q = 0$, és a dir, llevat que no hi hagi cap retard. En canvi, \tilde{R} és sempre una constant (en el supòsit de retards constants), i aquesta constant coincideix amb R sempre que els dos retards siguin iguals.

Si els retards no són constants, llavors els càlculs precedents no són exactes, però és d'esperar que encara serveixin com a primera aproximació.

En aquest sentit, sembla que podrien explicar l'anomalia que s'observa en el cas de Singapur. En efecte, en aquest país no s'està donant l'alta a un pacient fins després d'un període de seguretat que sembla ser bastant llarg, de l'ordre de setmanes (vegi's, per exemple, [18, punt 14]). Per tant, les altes s'estarien donant amb molt de retard respecte a la data real de la recuperació. Suposant que aquest retard q arribés a ser 15 dies superior a p , llavors amb una taxa de creixement $\delta = 0.2 \text{ dia}^{-1}$, que és un valor relativament normal, l'equació (17) dona una ratio $\tilde{R}/R = 20$, de l'ordre de magnitud del que s'observa en la figura 12.

D'altra banda, la bona concordància que s'observa en la part final de la gràfica corresponent a la Comunidad de Madrid (la primera de la figura 11) s'explicaria pel fet ja esmentat que des de fa un temps aquesta comunitat revisa sistemàticament l'historial a fi d'eliminar retards en la comptabilització de nous casos confirmats.

9 Conclusions

Tothom està d'acord que per a avaluar la gravetat de la situació en una epidèmia i prendre decisions al respecte és fonamental conèixer el valor del nombre de reproducció R i les seves variacions al llarg del temps. En particular, és crucial mantenir aquest paràmetre per sota de 1. Tanmateix, l'estimació del seu valor es pot fer de diverses maneres que fàcilment donen resultats diferents.

Aquestes diferències responen a diverses raons: (a) una diferent especificació del que es considera com a cas de la malaltia; (b) les pertorbacions causades pels casos descoberts a causa de campanyes generalitzades de proves de diagnòstic; (c) les hipòtesis relatives al mecanisme de propagació de la malaltia sobre les quals es basa el mètode d'estimació; i (d) la presència de retards entre els fets (casos nous, altes, morts) i la seva comptabilització.

En particular, les estimacions basades en el model SIR es veuen fàcilment afectades pels retards, especialment si aquests retards són variables i/o depenen del tipus d'esdeveniment.

Per tal de millorar les estimacions (encara que no es basin en el model SIR) és molt convenient eliminar aquests retards en la mesura que sigui possible. I a posteriori també resulta útil depurar les dades reassignant cada fet a la data en què va tenir lloc, tal com fan darrerament a la Comunidad de Madrid.

En qualsevol cas, cal ser conscients de l'existència d'aquests retards i tenir-los ben en compte a l'hora de dissenyar estratègies contra l'epidèmia.

Nota afegida el 20 de maig de 2020.

El dia 18 de maig de 2020 el Departament de Salut va corregir un error que hi havia en les seves dades, la qual cosa va comportar que el nombre acumulat de casos positius disminuís en 3 293 unitats respecte al dia anterior [12, Comunicat, 18 de maig de 2020]. Dos dies després ha publicat la sèrie històrica revisada de casos positius i de defuncions [12, Comunicat, 20 de maig de 2020]; de moment no s'ha publicat, però, la corresponent sèrie històrica d'altes, sense la qual no podem aplicar l'estimació de R basada en el model SIR.

D'altra banda, en les dades que publica el Ministerio de Sanidad, el dia 10 de maig el nombre de casos nous a Catalunya va ser anormalment alt a

causa de retards de notificació. Concretament, aquell dia es van notificar 2711 casos nous; posteriorment les dades d'aquells dies han sofert algunes revisions, però hi segueix havent un increment anormalment alt a causa de retards. Això —i la mitjana mòbil que fem de 7 dies— significa que els valors de R que es mostren al final de les figures 3 i 11 són més alts del compte.

Agraïment. Moltes gràcies per la seva col·laboració i comentaris a Carles Barril, Àngel Calsina, Sílvia Cuadrado, Mercè Farré, Clara Prats, Pere Puig i Jordi Ripoll.

Referències

- [1] Fred Brauer, Carlos Castillo-Chávez, 2012. *Mathematical Models in Population Biology and Epidemiology* (2nd ed). Springer.
- [2] Abel Cintrón-Arias, Carlos Castillo-Chávez, Luis M. A. Bettencourt, Alun L. Lloyd, H. T. Banks, 2009. The estimation of the effective reproductive number from disease outbreak data. *Mathematical Biosciences and Engineering*, 6: 262–282. [↗](#)
- [3] Anne Cori, Neil M. Ferguson, Christoph Fraser, Simon Cauchemez, 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178: 1505–1512. [↗](#)
- [4] Odo Diekmann, Hans Heesterbeek, Tom Britton, 2013. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton Univ. Press.
- [5] Pauline van den Driessche, 2017. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2: 288–303. [↗](#)
- [6] Minah Park, Alex R. Cook, Jue Tao Lim, Yinxiaohe Sun, Borame L. Dickens, 2020. A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine*, 9: art.967. [↗](#)
- [7] J. Wallinga, M. Lipsitch, 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274: 599–604. [↗](#)

Les referències que segueixen inclouen les fonts de dades i les primeres anàlisis d'aquestes. Noti's que moltes d'aquestes referències tenen un contingut variable. Si no diem una altra cosa, hi hem accedit el dia 15 de maig de 2020.

- [8] Borja Andrino, Daniele Grasso, Kiko Llaneras: Así evoluciona la curva del coronavirus en España y en cada autonomía. *El País*. [↗](#)

- [9] Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine: Temporal variation in transmission during the COVID-19 outbreak. [↗](#)
- [10] Computational Biology and Complex Systems Research Group, UPC: COVID-19. [↗](#)
- [11] FT Visual & Data Journalism team: *Financial Times*. [↗](#)
- [12] Generalitat de Catalunya, Departament de Salut: Comunicats. *Canal Salut, Notes de premsa*. [↗](#)
- [13] Gobierno de España, Ministerio de Sanidad: Enfermedad por nuevo coronavirus, COVID-19. Situación actual. Actualización 84 (23 abril 2020). [↗](#)
- [14] Instituto de Salud Carlos III: COVID-19 en España. [↗](#)
- [15] Instituto de Salud Carlos III: Informes sobre la situación de COVID-19 en España. [↗](#)
- [16] Instituto de Salud Carlos III: Evolución diaria de la pandemia de COVID-19 en España (1 mayo 2020). [↗](#)
- [17] Singapore. Ministry of Health: Press Releases. [↗](#)
- [18] Singapore. Ministry of Health: Ministerial statement by Mr. Gan Kim Yong, Minister for Health, at Parliament, on the second update on whole-of-government response to COVID-19, 4th May 2020. [↗](#)
- [19] SUTD Data-Driven Innovation Lab: Predictive Monitoring of COVID-19. [↗](#)