

Đánh giá nguy cơ, đánh giá diễn biến, và dự báo cho dịch COVID-19

**Xêmina Xác suất và thống kê, Viện Toán học
26/08/2021**

Hà Thành Trung
Independent researcher
TS. Toán thống kê, UF
Postdoc Y tế cộng đồng, UCF

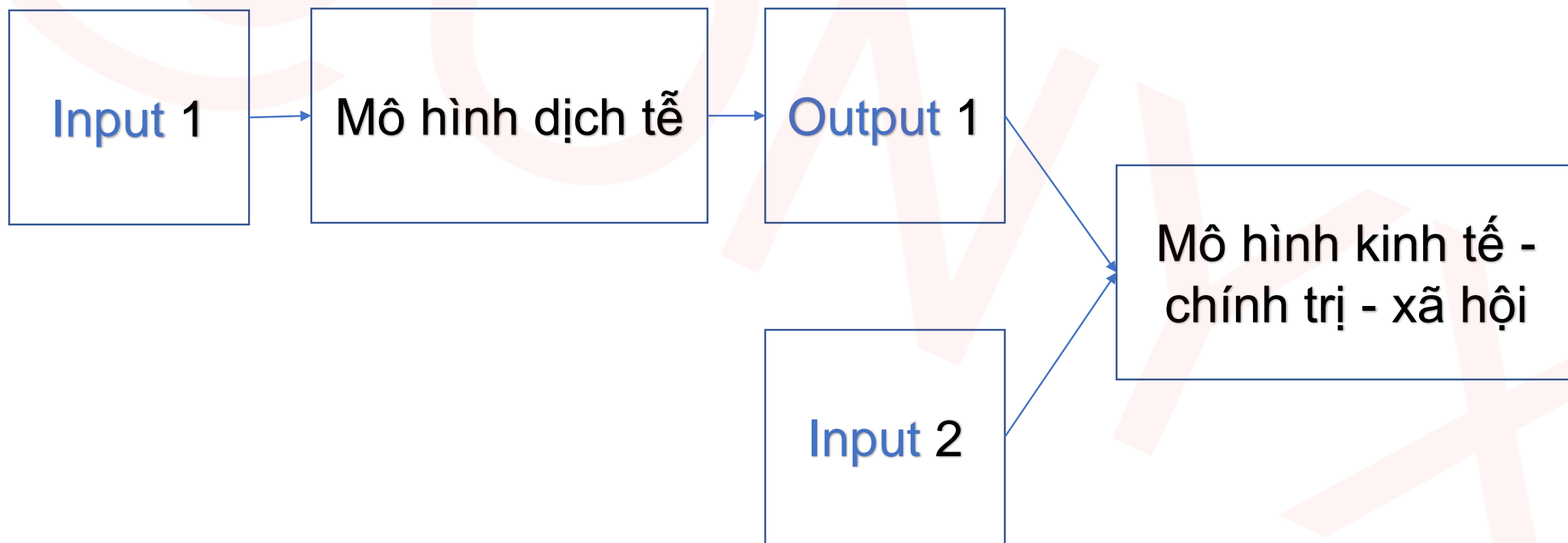
Nguyễn Khương Tuấn
Inventor
Nhà sáng lập Onyx, STECH

Nội dung

- Vai trò của mô hình dịch tễ
- Literature review
- Phần 1: Mô phỏng
- Phần 2: Ước lượng
- Phần 3: Dự báo
- Sóng siêu lây nhiễm
- Thang đo nguy cơ

Vai trò của mô hình dịch tễ

- Gồm 3 phần: mô phỏng (phần 1), ước lượng (phần 2), và dự báo (phần 3)



Literature review - Thế giới

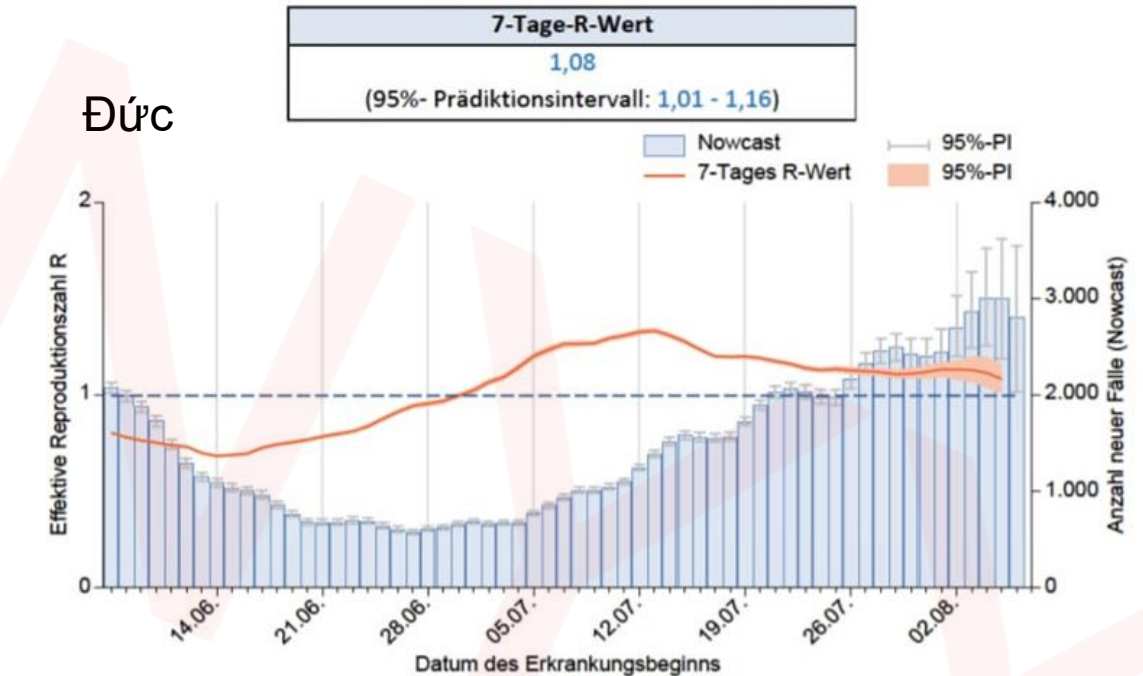
Mô hình dựa trên **2 tham số quan trọng** là **hệ số lây nhiễm thực** (effective reproduction number) và **thời gian nối tiếp** (serial interval) được sử dụng bởi Chính phủ:

Đức:

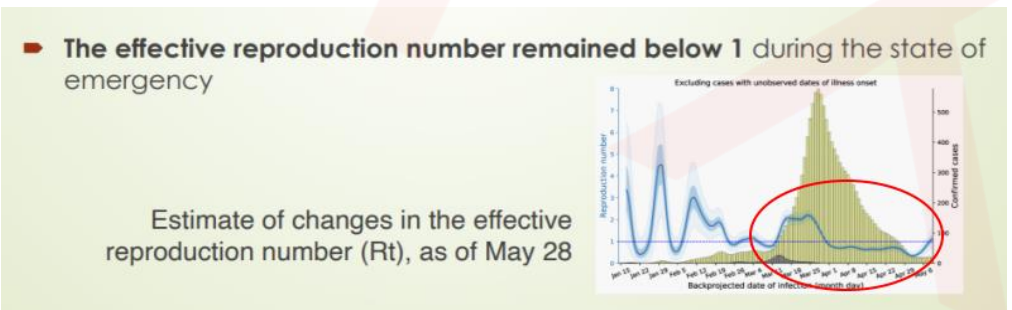
- Tính hệ số lây nhiễm thực hàng ngày
- Điều khiển nó ở mức gần 1
- Thời gian áp dụng: Từ ít nhất 04/20 đến nay

Một số Chính phủ/tổ chức khác:

- Anh (gov.uk)
- Ấn Độ
- CovidActNow (Mỹ)
- Nhật Bản
- Hồng Kông, Na Uy, và Tây Ban Nha



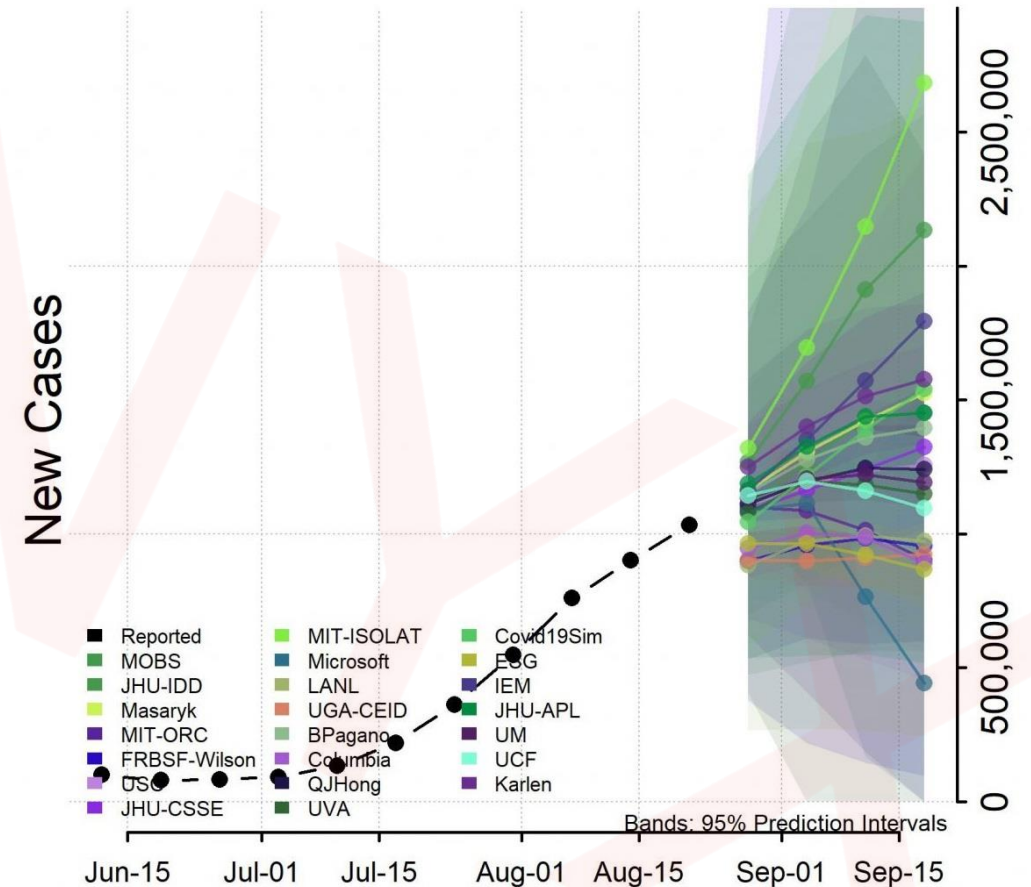
Nhật Bản



Mỹ - 102 mô hình

- Dữ liệu công khai (CDC Mỹ cung cấp)
- Luật tham gia rõ ràng, trách nhiệm rõ ràng
- Văn bản công khai (tương tự với Chính phủ Anh và Đức)
- Ai cũng có thể tham gia ([102 mô hình](#), một số ngừng cập nhật), công khai phương pháp
- Mô hình tin cậy nhất là sự tổng hợp của hơn 20 mô hình ([ensemble](#)).
- Từng có một cá nhân làm ra mô hình chính xác nhất ([Youyang Gu](#))

National Forecast



[CDC Mỹ](#) - mô hình

Việt Nam

Một số tổ chức/nhóm/cá nhân chúng tôi được biết (xin lỗi nếu chúng tôi không tìm ra bạn):

1. [TS. Lê Anh](#) là người đầu tiên làm xong (28/03/20). Dùng một lớp các đường cong một đỉnh (2 tham số). Cách để tìm ra đường cong rất bài bản và thú vị. Ưu điểm: đơn giản, dự báo đỉnh khá chính xác khi dịch chỉ có 1 đỉnh, dữ liệu đủ tốt. Nhược điểm: thiếu phần 1 và phần 2; input không có các chủng vi-rút khác nhau; có thể bị đánh lừa bởi đỉnh địa phương (tù) dẫn đến tâm lý chủ quan; và không chỉ ra kịch bản xấu.
 2. [Viện Max-Planck](#), với TS. Đức là thành viên. Chúng tôi mới biết nhóm và chưa kịp xem kỹ.
 3. [Onyx](#). Trung phát triển tới 05/20. Trung kết hợp Tuần để thêm smoothing spline, sóng siêu âm nhiễu, cũng như tích hợp công nghệ chống giả iSeal từ 07/21.
 4. [PGS. Truyền](#) công bố trên FB khoảng 09/2020. Dùng tổng tuyến tính hai/ba đường cong nên linh hoạt hơn nhưng cũng nhiều tham số hơn mô hình của TS. Lê Anh.
 5. 31/05/21, Ban chỉ đạo Quốc gia phòng, chống dịch COVID-19 có [Quyết định 2686/QĐ-BCĐQG](#) và [bản đồ nguy cơ](#) (đang thử nghiệm).
- Ngoài ra rất nhiều nhóm khác nhưng không rõ họ có áp dụng mô hình cho dữ liệu hiện nay như [5F](#), nhóm sử dụng mô hình của [Quỹ Bill-Melinda Gates](#), [ĐH Duy Tân](#), [Oucru](#), và Tech4Covid (mô hình của [TS. Quoc Tran](#)).

Phần 1 – Mô phỏng giai đoạn đầu

- Vấn đề: khi chưa có dữ liệu thì ta có thể làm gì?
- Giải pháp: Lịch sử. Xem tất cả các nước.
- Điều kiện cần: Khả thi khi một chủng COVID-19 đã xảy ra ở nước khác. **Hệ số lây nhiễm cơ bản** R_0 và **thời gian nối tiếp trung bình** T đã được ước lượng.
- Ý nghĩa: **1 ca bệnh tạo ra R_0 ca bệnh trong khoảng thời gian T** (tương tự GDP được tính trong 1 năm).
- Trong giai đoạn đầu nT ngày, khi $R_0 > 1$, 1 cá thể bệnh sẽ sinh ra tổng số cá thể bệnh (trung bình):

$$\frac{R_0^{n+1} - 1}{R_0 - 1}$$

- R_0 cho vùng thưa dân cư sẽ nhỏ hơn R_0 cho thành phố.
- R_0 là không đủ nếu thiếu T . Cặp ($R_0 = 5, T = 4$) sẽ khiến số ca tăng nhanh hơn cặp ($R_0 = 25, T = 8$).

$$T = 4$$

Chủng COVID-19	R_0	Tổng số ca sau 20 ngày
SARS-CoV-2	2	63
Alpha	3.4	643
Delta	5	3906

Chính phủ Anh có [báo cáo mỗi 2 tuần](#) về các chủng

Phần 1 – Phát hiện ổ dịch

- Vấn đề: ước lượng kích thước ổ khi phát hiện ổ dịch.
- Input: Ngoài R_0 và T , ta cần thêm 2 tham số: xác suất phát hiện ca bệnh p và thời gian trung bình từ khi mắc bệnh đến khi được phát hiện τ .
- Thuật toán:
 - Bước 1: Với xác suất 95%, tính số ca m tối thiểu để phát hiện ra ổ.
$$(1 - p)^m < 5\%$$
 - Bước 2: Quy đổi m ra số ngày.
- Ví dụ với $p = 10\%$, $\tau = 7$:

Chủng COVID-19	R_0	Số ngày để có ít nhất 29 ca	Số ngày cho tới khi đi viện	Tổng số ca khi phát hiện ổ
SARS-CoV-2	2	16	23	63 đến 127
Alpha	3.4	12	19	188 đến 643
Delta	5	8	15	156 đến 781

- Ví dụ: Áp dụng ở Đà Nẵng vào 07/2020.

Phần 2 – Đặt vấn đề

- Chúng tôi chỉ có trong tay:
 - Input: số ca được phát hiện mỗi ngày (reported)
 - Có các ca xâm nhập
- Cần giải quyết:
 - Có các biện pháp can thiệp như giãn cách, khẩu trang, cách ly .v.v. tại các thời điểm khác nhau
- Ta sẽ xem xét một phương pháp cân bằng giữa sự **hợp lý** và tính **giải được**:
 - Xuất phát từ các khái niệm cơ bản để lựa chọn mô hình **time-since-infection** (sử dụng mô hình SIR để mô tả nó).
 - Bám sát câu hỏi để tìm loại **reproduction number** phù hợp
 - Những vấn đề khi áp dụng

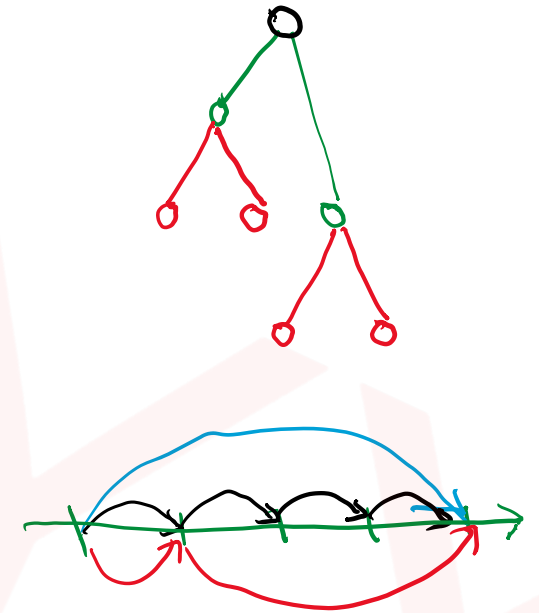
Nhìn nhận vấn đề

Vấn đề có sự tham gia của 3 yếu tố cơ bản:

- Môi trường,
 - Đối tượng,
 - Quá trình tăng/giảm số lượng của đối tượng.
- Môi trường là nơi đối tượng có thể tồn tại (người, dơi, .v.v.). Nó thường **hữu hạn**. Ở đây giới hạn trong loài người.
 - Đối tượng có thể gồm nhiều loại loài/hạt/.v.v. cùng một lúc như trong hệ sinh thái. Ở đây là vi-rút.
 - Trường hợp của ta, quá trình tăng/giảm vi-rút được đơn giản hóa thành quá trình tăng/giảm số cá thể bệnh. Tuy nhiên ta vẫn mô hình hóa nồng độ vi-rút trong cá thể bệnh (**viral load**).
 - Quá trình tăng giảm số ca bệnh được tính qua các hệ số lây nhiễm.

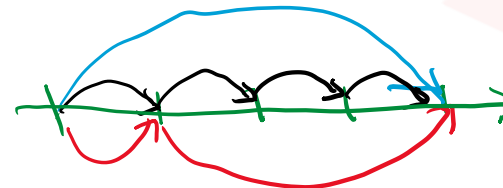
Effective reproduction number

- **Effective reproduction number $R_e(t)$** : the average number of secondary infections caused by a typical primary infection at time t [21]
- Đặc điểm:
 - Liên hệ giữa hai **thế hệ** (primary, secondary). Hệ quả là tập cho **primary** và tập cho **secondary tách rời**.
 - Chỉ lây **một** lần duy nhất (hai thế hệ liên tiếp)
 - Chỉ dễ hiểu khi các cá thể có cùng R_e (**đồng nhất**). Khi đó để tính R_e , ta chỉ cần số ca trong mỗi thế hệ.
- Một số vấn đề:
 - Có nhiều thế hệ tại cùng một thời điểm.
 - Nếu không có quy luật cho $R_e(t)$, thì ta có n biến trong n ngày. Dữ liệu đầu vào chỉ có n số cho n ngày. Do đó $R_e(t)$ phải là biến duy nhất thì bài toán mới **giải được**. Ta xét một mô hình thỏa mãn điều này.



Ba giả thiết cơ bản

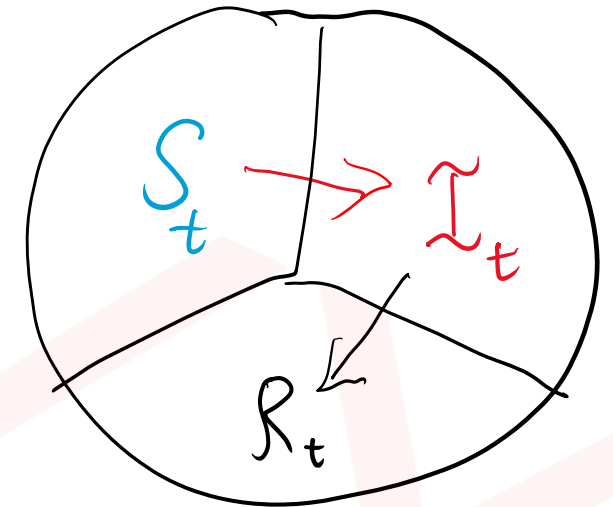
- (Rời rạc) Do dữ liệu là số ca mới mắc mỗi ngày, giả thiết thời gian là ngày ($t \in \mathbb{N}$). Điều này giúp đơn giản hóa mô hình.
- (Độc lập) Để giải quyết vấn đề có nhiều thể hệ tại một thời điểm, giả thiết quá trình truyền bệnh của mỗi cá thể là độc lập. Như vậy quá trình truyền bệnh của một cá thể phụ thuộc thời điểm nó mới mắc bệnh và quá trình truyền bệnh của nó, chứ không phụ thuộc vào thể hệ của nó. Từ đó ta có thể gán thể hệ cho các ca bệnh và tìm các đại lượng tương tự R_e để đánh giá dịch.
- (Đồng nhất) Để vấn đề có thể giải được với dữ liệu chỉ là số ca mới mắc mỗi ngày, giả thiết quá trình truyền bệnh là như nhau cho các cá thể mới mắc bệnh tại cùng một thời điểm. Như vậy sự loại bỏ khỏi quá trình truyền bệnh cũng đồng nhất. Nếu không có tính đồng nhất, thì cần thêm thông tin về từng nhóm và sự truyền bệnh giữa các nhóm.
- Ta sẽ tìm cách dựng lại quá trình bệnh lan truyền từ dữ liệu (ở mức đủ để trả lời câu hỏi, chứ không biết ai lây cho ai).



SIR (thời gian rời rạc)

Nhắc lại mô hình SIR [11]:

- **Giả thiết 1:** thời gian là ngày ($t \in \mathbb{N}$) và sự kiện gặp gỡ cũng như lây bệnh chỉ xảy ra tại các thời điểm cách nhau đúng 1 ngày
- Tập vật chủ được chia thành 3 tập thay đổi theo thời gian:
 - S_t : (**susceptible**) lành không có miễn dịch, có thể nhiễm bệnh
 - I_t : (**infective**) bệnh, có thể lây bệnh cho cá thể trong S_t
 - R_t : (**removed**) không liên quan đến quá trình truyền bệnh (ví dụ như cách ly (**isolation**), có miễn dịch, khỏi bệnh/chết)
- S_t, I_t : lần lượt là số cá thể trong S_t và I_t
- **Giả thiết 2:** Tổng số cá thể $N > 0$ không đổi



- **Giả thiết 3:** Một cá thể gặp $m > 0$ cá thể bất kì 1 ngày. Các sự kiện gặp gỡ là độc lập.

Để lây bệnh cần cá thể trong \mathcal{J}_t gặp cá thể trong \mathcal{S}_t ngày $t + 1$.
 Một cá thể trong \mathcal{J}_t gặp $m \frac{S_t}{N}$ cá thể trong \mathcal{S}_t trong ngày.

- **Giả thiết 4a:** Tỷ lệ lây bệnh là $0 < p < 1$ cho mỗi gặp gỡ giữa \mathcal{J}_t và \mathcal{S}_t .

Gọi \mathcal{V}_t là tập cá thể mới mắc bệnh tại t . V_t là số cá thể trong \mathcal{V}_t .

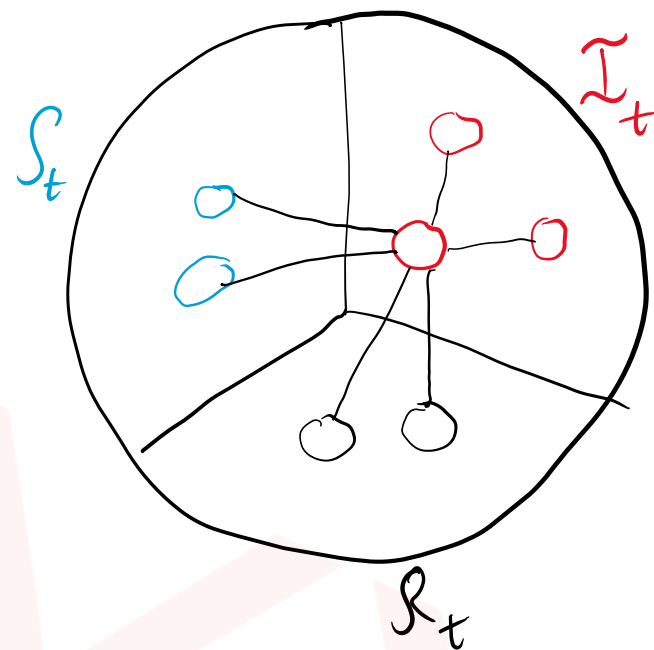
$V_{t+1} = (\text{số cá thể } \mathcal{J}_t) * (\text{số cá thể } \mathcal{S}_t \text{ mà 1 cá thể } \mathcal{J}_t \text{ gặp}) * (\text{tỷ lệ lây})$

$$= I_t \left(m \frac{S_t}{N} \right) p = \frac{mp}{N} S_t I_t =: \beta S_t I_t, \text{ với } \beta = \frac{mp}{N}.$$

Ta có

$$V_{t+1} = \beta S_t I_t$$

- βS_t là tốc độ truyền bệnh tại thời điểm t
- Chú ý: ta xét m và p thay đổi theo thời gian sau này. Ta không xét phương trình thứ 2 của SIR vì ta không dùng nó ở đây.



$$m \frac{1}{3} = 2 \frac{S}{N}$$

$$V_t = \beta S_t I_t = \frac{mp}{N} S_t I_t$$

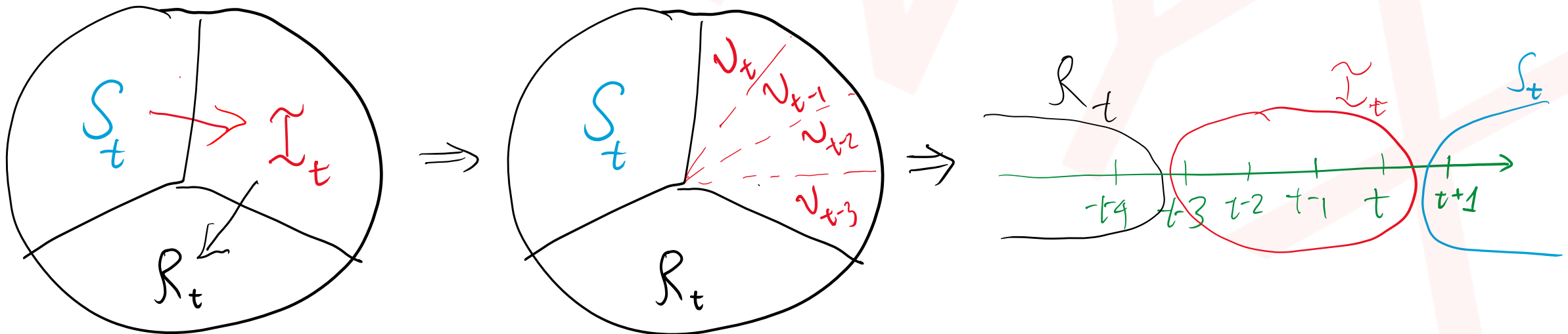
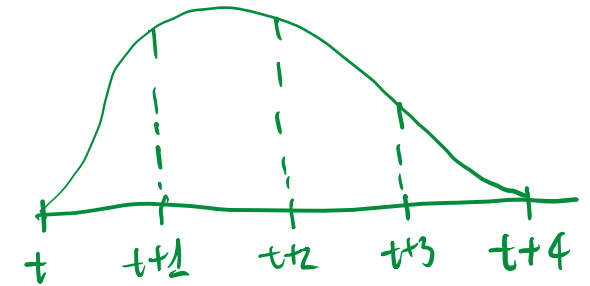
Các cách làm giảm số ca mắc mới (phù hợp kinh nghiệm thực tế):

- Giảm tỷ lệ lây bệnh p : khẩu trang, rửa tay, .v.v.
- Giảm số gặp gỡ m : giãn cách xã hội, quãng đường di chuyển, .v.v.
- Giảm số cá thể không miễn dịch S_t : vắc-xin
- Giảm số ca bệnh I_t : cách ly, nhập viện

Mô hình time-since-infection

Cả SIR và **time-since-infection** với tham số không đổi theo thời gian là các trường hợp riêng của Kermack-McKendrick (KM) model. Các giả thiết 1-3 vẫn như SIR [4, 11]. **Time-since-infection** cho phép nhiều tốc độ lây thay vì chỉ một trong SIR.

- **Giả thiết 4b**: Khi 1 cá thể **mới** mắc bệnh tại $t \in \mathbb{N}$, nó chỉ lây cho cá thể khác trong T ngày tiếp theo. Tỷ lệ lây bệnh là p_k cho mỗi gặp gỡ sau k ngày (gặp gỡ tại thời điểm $t + k$ với $k = 1, \dots, T$), trong đó $\sum p_k > 0$. **Để tiện cho trình bày, chọn $T = 4$.**
- **Giả thiết 5**: Không cá thể bệnh nào bị loại bỏ trong vòng T ngày kể từ khi mới mắc bệnh.
- S_t, I_t , và R_t được biểu diễn lại trên trục thời gian. Chú ý $I_t = V_t + \dots + V_{t-3}$.

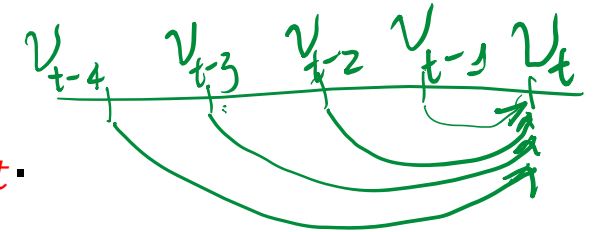


- Từ nay về sau chỉ xét $t \geq T$. (Chú ý ta thay $t + 1$ bởi t so với trang trước).
- Tương tự SIR, tại thời điểm t ta có $\beta_k = mp_k/N$ và \mathcal{V}_{t-k} lây cho $(\beta_k S_{t-1})V_{t-k}$ cá thể tại thời điểm t ($k = 1, \dots, 4$). Do đó

$$V_t = \beta_1 S_{t-1} V_{t-1} + \dots + \beta_4 S_{t-1} V_{t-4} = S_{t-1} (\beta_1 V_{t-1} + \dots + \beta_4 V_{t-4}).$$

- Chuẩn hóa $w_k = \beta_k / (\sum \beta_j)$ thì $\sum w_k = 1$ và

$$V_t = \underbrace{[S_{t-1} \sum \beta_k]}_{R_t} \underbrace{(w_1 V_{t-1} + \dots + w_4 V_{t-4})}_{\Lambda_t} =: R_t \Lambda_t.$$



- Tốc độ lây được biểu diễn lại qua R và w thay vì S và β do

$$\beta_k S_{t-1} = S_{t-1} [w_k \sum \beta_j] = w_k R_t.$$

\Rightarrow Một cá thể mới mắc bệnh tại thời điểm $t - k$ lây cho $w_k R_t$ cá thể tại thời điểm t với $\sum w_k = 1$. R được gọi là **instaneous reproduction number**.

- Nếu biết w , từ V cho $t \leq t_0$, ta tính ra R cho $T \leq t \leq t_0$. Từ đó ta có thể **dựng lại quá trình lan truyền bệnh** cho $T \leq t \leq t_0$ (Theo nghĩa không biết ai lây cho ai, nhưng biết như 3 người trong \mathcal{V}_s lây cho 6 người trong \mathcal{V}_t).

Xác định **đại lượng** để đánh giá diễn biến dịch

- R (**instaneous reproduction number**) không phù hợp vì nó chỉ nói lên 1 thay vì T ngày.
- **Case reproduction number** $R_c(t)$ quan tâm một cá thể trong \mathcal{V}_t sẽ lây cho bao nhiêu cá thể qua một lần lây.

$$R_c(t) := R_{t+1}w_1 + \dots + R_{t+4}w_4.$$

R_c phù hợp nhưng cần biết w .

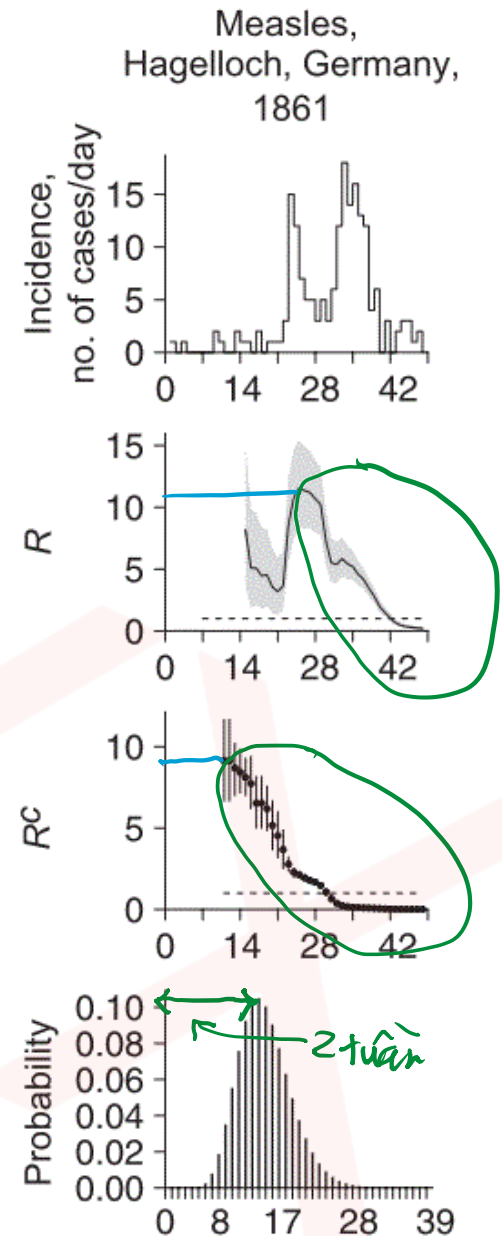
- Hệ số lây nhiễm R_T ($I_t > 0$)

$$R_T(t) = \frac{I_{t+T}}{I_t}.$$

Chúng tôi chọn R_T (dịch chuyển 4 ngày so với [17]) vì nó không cần w , dễ hình dung, và dễ suy luận.

Ví dụ thực tế về R và R_c

- Hình bên phải từ [1]. Đường xanh da trời cho thấy đỉnh R_c thấp hơn. Vòng tròn xanh lá cây cho thấy độ trễ giữa R và R_c . Độ trễ là khoảng 2 tuần. Nó xấp xỉ vị trí đỉnh của w .



Tham số thay đổi theo thời gian

Time-since-infection cho phép tham số thay đổi [4]. Các biện pháp can thiệp làm thay đổi số gặp gỡ m_t , xác suất lây $p_{k,t}$, và thời gian truyền bệnh T_t . N_t và S_t cũng có thể thay đổi. Chú ý rằng ta chỉ cần T và w không đổi. Cách ly sẽ làm T nhỏ đi và ta thường cố định giá trị của nó. **Vậy còn giả thiết w không đổi thì sao?** Vì w đã được chuẩn hóa $\sum w_{k,t} = 1$ nên chỉ cần tỷ lệ $w_{k,t}:w_{1,t}$ không đổi với $k = 2, \dots, T$. (Để đơn giản giả sử $w_{1,t}, \beta_{1,t}, p_{1,t} > 0$.)

- Vì $\beta_{k,t} = m_t p_{k,t} / N$ nên $\beta_{k,t}:\beta_{1,t} = p_{k,t}:p_{1,t}$. Chú ý $\beta_{k,t}:\beta_{1,t} = w_{k,t}:w_{1,t}$, do đó

$$w_{k,t}:w_{1,t} = p_{k,t}:p_{1,t}$$

không phụ thuộc m_t . Nếu các biện pháp can thiệp thay đổi các $p_{t,k}$ theo cùng tỷ lệ, thì $p_{k,t}:p_{1,t}$ không đổi. Từ đó $w_{k,t}:w_{1,t}$ cũng không đổi.

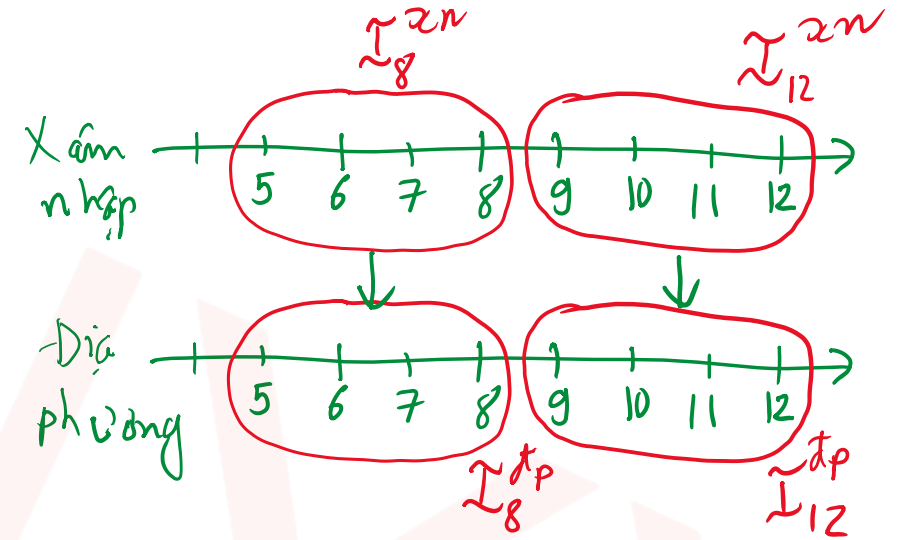
- Nếu coi như $p_{k,t}:p_{1,t}$ chỉ phụ thuộc nồng độ virus, thì giả thiết w không phụ thuộc giãn cách và khẩu trang là chấp nhận được.
- Tính đồng nhất vẫn giữ nguyên do cùng tốc độ truyền bệnh cho các cá thể trong \mathcal{V}_t .

Xâm nhập

- Khi không có xâm nhập, nhắc lại rằng $R_T(t = 8) = \frac{I_{12}}{I_8} = \frac{V_9 + \dots + V_{12}}{V_5 + \dots + V_8}$ (với $I_8 > 0$)

Khi có xâm nhập, thay J_t bởi $J_t^{\text{đp}}$. Gọi tập các ca xâm nhập trong khoảng thời gian từ $t - T + 1 = 5$ đến $t = 8$ là $J_{t=8}^{\text{xn}}$ (hình dung ta có thể gộp các khối J_t^{xn} và J_t^{ng}). Vì $J_{t=12}^{\text{xn}}$ không được tạo ra bởi $J_8^{\text{đp}}$ và J_8^{xn} , ta nên loại bỏ chúng và các chuỗi lây nhiễm được tạo bởi chúng ra khỏi tử số. Tuy nhiên ta không có thông tin về chuỗi lây nhiễm được tạo ra vào tử số (trừ chính các ca xâm nhập). Khi đó ($I_8^{\text{đp}} + I_8^{\text{xn}} > 0$)

$$\hat{R}_T(t = 8) = \frac{I_{12}^{\text{đp}}}{I_8^{\text{đp}} + I_8^{\text{xn}}}$$



- Không xử lý xâm nhập càng gây **overestimate**.
- Ta đơn giản hóa vấn đề bằng cách giả sử ca xâm nhập có cùng w . Thực tế ca xâm nhập có thể bị cách ly sớm hơn nên có $T_{\text{xâm nhập}}$ khác T . Ta không giải quyết vấn đề đó ở đây (vấn đề không đồng nhất).
- [Thompson et al. 2019] mở rộng phương pháp của [Cori et al. 2013] bằng cách bỏ các ca xâm nhập khỏi tử số.

Ngẫu nhiên hóa

- Khi dữ liệu đầu vào V_t không chính xác (**missing data**, độ trễ khi thu thập dữ liệu, .v.v.), ta có thể giả thiết một phân phối cho toàn bộ dữ liệu ($V_0, V_1, \dots, V_t, \dots$).
- Có rất nhiều cách khác nhau để ngẫu nhiên hóa. Ví dụ như tùy vào việc thu thập dữ liệu [17], hoặc sử dụng một phân phối phổ biến như quá trình **Poisson** cho dữ liệu đếm [1].
- Nếu giả thiết không phù hợp/**robust**, thì **confidence/credible interval** chỉ tạo ảo giác và có thể phản tác dụng.
- Do quan hệ giữa V và R_T là 1-1, và chúng tôi không có thông tin nào về thu thập dữ liệu, chúng tôi coi như R_T là ngẫu nhiên. Tìm hiểu các nước có dữ liệu tốt như Đức, chúng tôi thấy quỹ đạo R_T khá ổn định, và các nhà nghiên cứu chỉ ra rằng R_T biến đổi độ dốc khi hành vi đám đông thay đổi. Từ đó chúng tôi dùng hồi qui phi tham số (smoothing spline) để làm trơn R_T . Chạy một số mô phỏng đơn giản, chúng tôi cũng quan sát thấy làm trơn R_T là phù hợp.

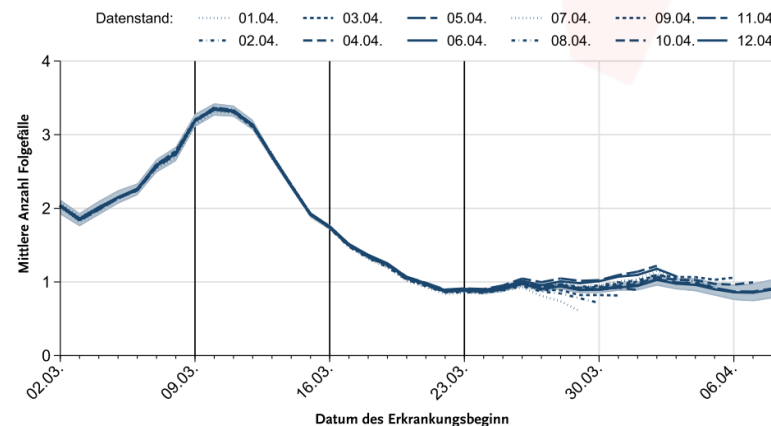


Abb. 5 | Vergleich der Schätzung der effektiven Reproduktionszahl R für eine angenommene Generationszeit von 4 Tagen zu unterschiedlichen Datenständen. Die schwarz hervorgehobenen vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

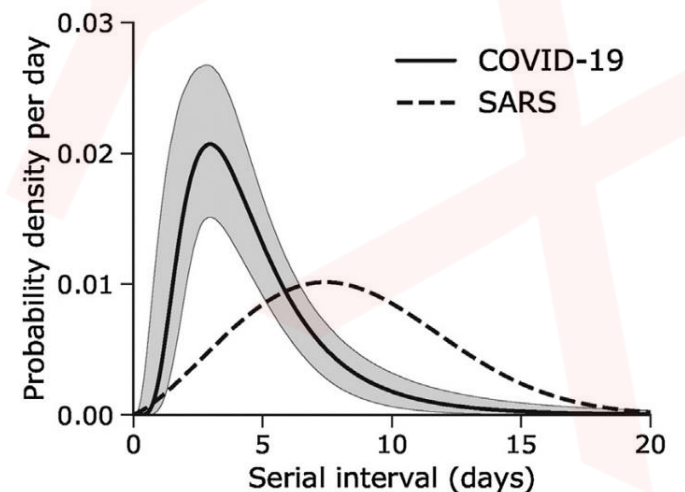
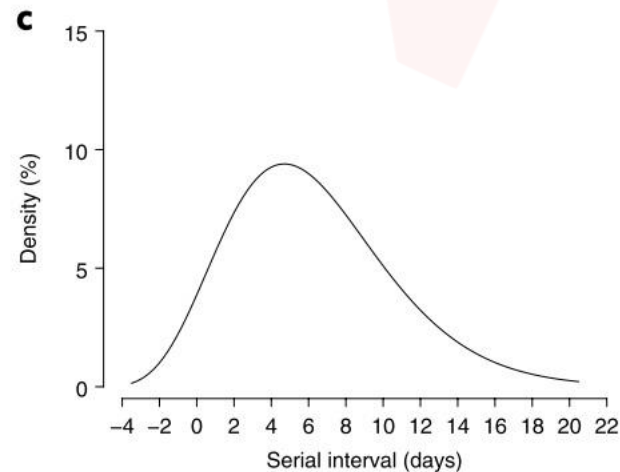
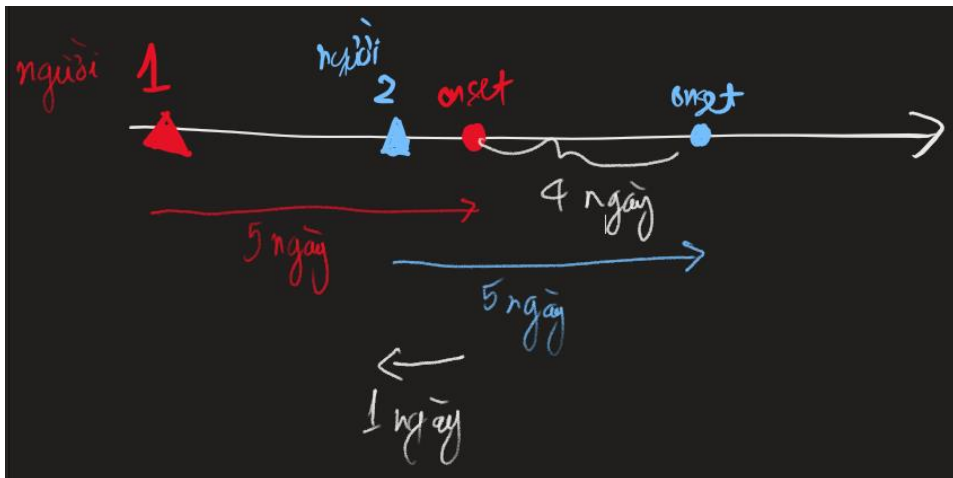
Ứng dụng

- Với một cá thể ta có các thời điểm

Infection of a virus < onset of symptoms < removal

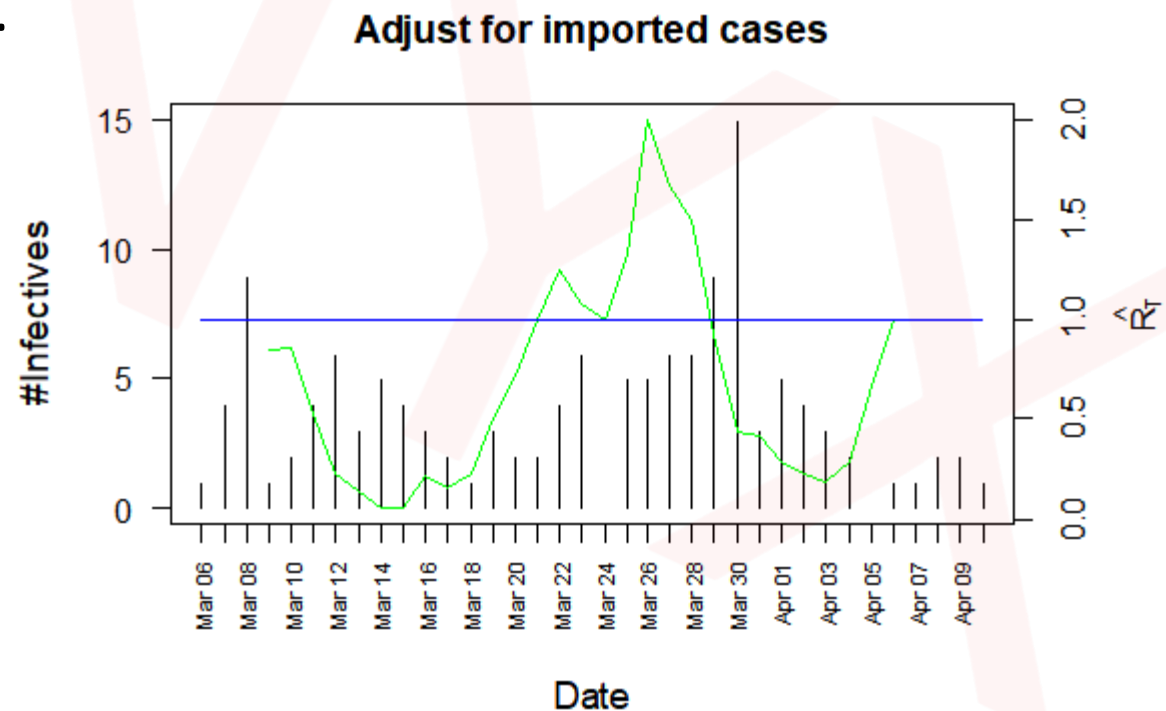
Incubation time: khoảng thời gian giữa **infection of a virus** và **onset**

- Khi hai cá thể lây cho nhau, ta có các khoảng thời gian
 - **generation time:** giữa 2 **infections**
 - **serial interval (SI):** giữa 2 **onsets**, có thể âm (hình dưới) [8, 13]
- Thực tế ta chỉ quan sát được **onset of symptoms**. Nên V_t sẽ đếm **onset**, và w sẽ cho **SI**.
- Vì các hệ số lây nhiễm được tính cho **onset** và SI, ta cần chuyển đổi nó sang hệ số lây nhiễm **thật** thông qua các giả thiết khác (như **incubation time** độc lập SI, .v.v. [1]).



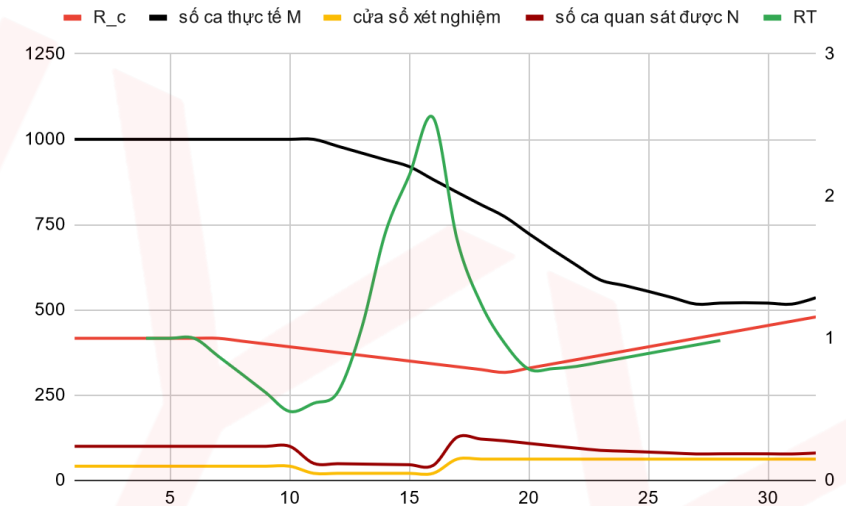
Đợt dịch đầu ở Việt Nam

- Ước lượng \hat{R}_T khi điều chỉnh cho xâm nhập
- Nếu mọi giả thiết là hợp lý (mô hình, dữ liệu), thì Việt Nam hoàn toàn kiểm soát được dịch. Đồ thị \hat{R}_T với 4 đỉnh dương như liên hệ tới các cụm dịch Trúc Bạch-Bình Thuận, rồi Buddha, rồi Bạch Mai-Trường Sinh, và cuối cùng là Mê Linh. Như vậy biện pháp **contact tracing** có tác dụng kiểm soát dịch thành các cụm. Các biện pháp xã hội không thể hiện rõ rệt do có thể được áp dụng trước 06/03/20 và/hoặc **contact tracing** có tác dụng lẫn át.
- Khi \hat{R}_T chỉ phản ánh các cụm dịch nhỏ thì dùng \hat{R}_T để làm dự báo là không phù hợp
- Phương pháp phù hợp dữ liệu nhỏ nếu dữ liệu tốt



Xem lại một số giả thiết

- SI dương: khi SI âm thì một cá thể trong J_{t+T} trong khi nó nên trong J_t và ngược lại. Điều này dễ xảy ra tại các thời điểm gần nhau giữa 2 tập. Khi tính R_T ta nhóm T ngày lại. Điều này sẽ làm giảm bớt hiệu ứng của SI âm.
- Không triệu chứng: Kể cả khi V_t được thu thập đầy đủ cho các ca có triệu chứng, thì các ca không triệu chứng vẫn gây vấn đề. Ta cần một giả thiết về quy luật giữa không triệu chứng và có triệu chứng. Ví dụ nếu tỷ lệ giữa hai loại không thay đổi theo thời gian thì các phương pháp ở đây vẫn dùng được [1].
- Mở rộng điều trên thì có nghĩa rằng ta cần tỷ lệ giữa số ca được phát hiện và số ca thực tế ổn định. Điều này vốn phụ thuộc tỷ lệ đi viện (khá ổn định) và thuật toán xét nghiệm. Ta cần số xét nghiệm ổn định hàng ngày cho các ca không đi viện và xét nghiệm được thực hiện với cùng quy luật survey.
- Chúng tôi mô phỏng để xem xét việc thay đổi quy mô xét nghiệm ảnh hưởng R_T ra sao. Nếu số xét nghiệm thay đổi lần lượt 10% - 5% - 15%, thì R_T bị under-estimate, rồi lại over-estimate, sau đó quay lại quỹ đạo đúng. Đường spline sẽ giúp ích trong trường hợp này.



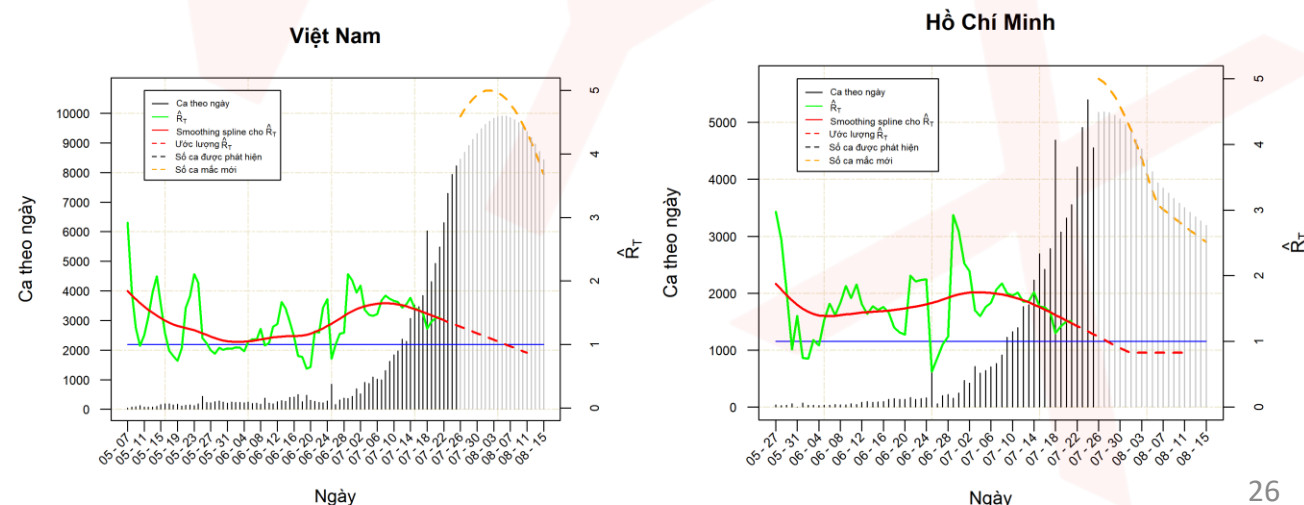
Phần 3 – Compartmental model

Giả sử trong tương lai **compartmental model** phù hợp và ta xấp xỉ được R_e của mô hình (chả hạn bằng R_c hay R_T , rồi hiệu chỉnh phù hợp bằng hồi quy). Chúng tôi chọn SEIR. Kí hiệu t_0 là điểm xuất phát.

- Nếu sử dụng **compartmental model** khác với nhiều **compartment** hơn thì cần ước lượng nhiều tham số hơn. Các thủ tục tương tự. Công thức cho R_e được tính qua **next-generation method** [3].

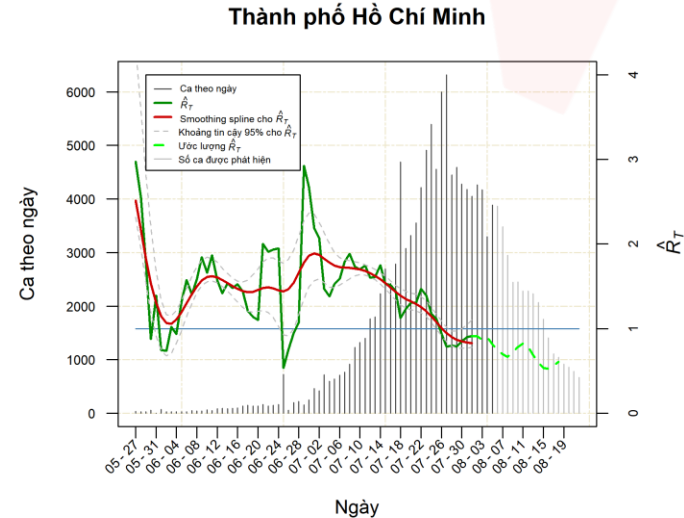
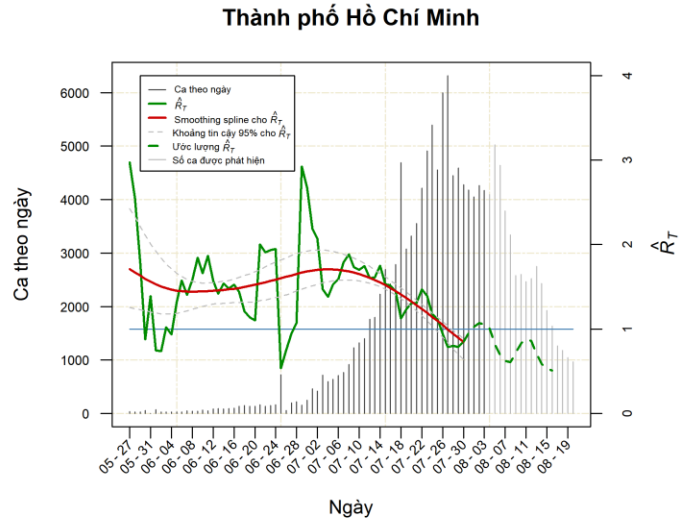
Tham số	Giá trị	Lý do
N: Tổng dân số Việt Nam	98 triệu	Xấp xỉ
E_0 : Số ca bệnh chưa được phát hiện hôm t_0	Tổng số ca phát hiện từ $t_0 + 1$ đến $t_0 + 4$	Đây là khoảng thời gian đa số những ca này được phát hiện
I_0 : Số ca bệnh được phát hiện trong vòng T ngày cho tới t_0	Tổng số ca phát hiện từ $t_0 - 3$ đến t_0	
Remove ₀ : Số ca không liên quan quá trình lây bệnh	1 triệu	Không ảnh hưởng quá nhiều tới mô hình
Gamma: Thời gian ủ bệnh trung bình	4 ngày	Thời gian ủ bệnh trung bình cho chủng Delta khoảng 4 ngày
Delta	4 ngày	T
R_e	Ước lượng của R_T hôm t_0	
Beta	Tính từ các tham số trên	

- Chúng tôi đưa ra một kịch bản dự báo khi xu hướng đường spline của R_T rõ ràng.
- Trong các trường hợp khác, chúng tôi chọn nhiều kịch bản dự báo.
- Chúng tôi không tính khoảng tin cậy vì khoảng tin cậy cho spline ngoài khoảng dữ liệu rất lớn.

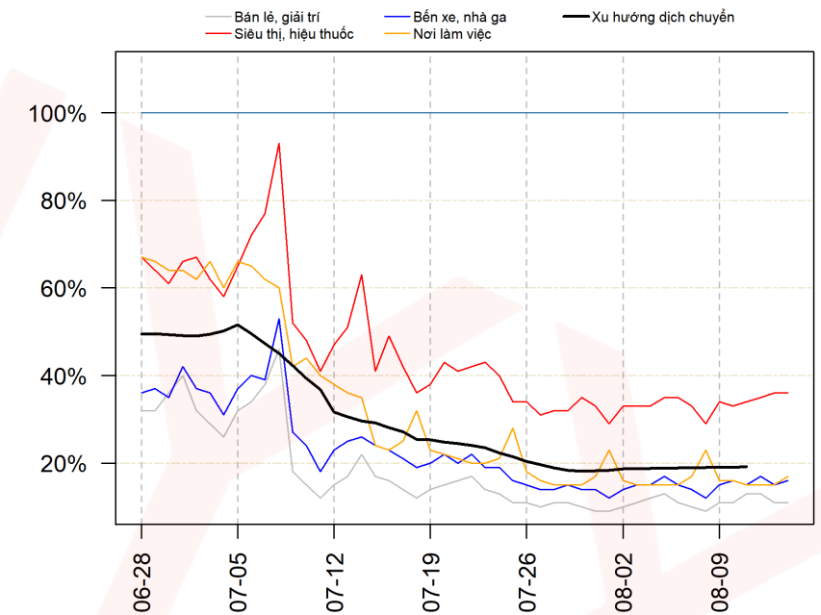


Sóng siêu lây nhiễm

- Sau khi dự báo 2 kịch bản cho Việt Nam ngày 18/07, chúng tôi quan sát thấy quỹ đạo dữ liệu lạ thường. Kết hợp với Google Mobility, chúng tôi cho rằng có thể có sự kiện siêu lây nhiễm (SKSLN) ở Thành phố Hồ Chí Minh (HCM) vào ngày 08/07 (trước khi thực hiện Chỉ thị 16) [6].
- Dữ liệu TP HCM phù hợp vì nó không có chu kỳ 7 ngày, hành vi tụ tập chỉ xảy ra đúng 1 ngày, và phong tỏa xảy ra ngay sau đó [16].
- Chúng tôi sử dụng cùng mô hình, và thêm giả thiết: SKSLN tạo ra thêm 1.618 ca vào ngày 18/07.



Xu hướng dịch chuyển ở TP Hồ Chí Minh

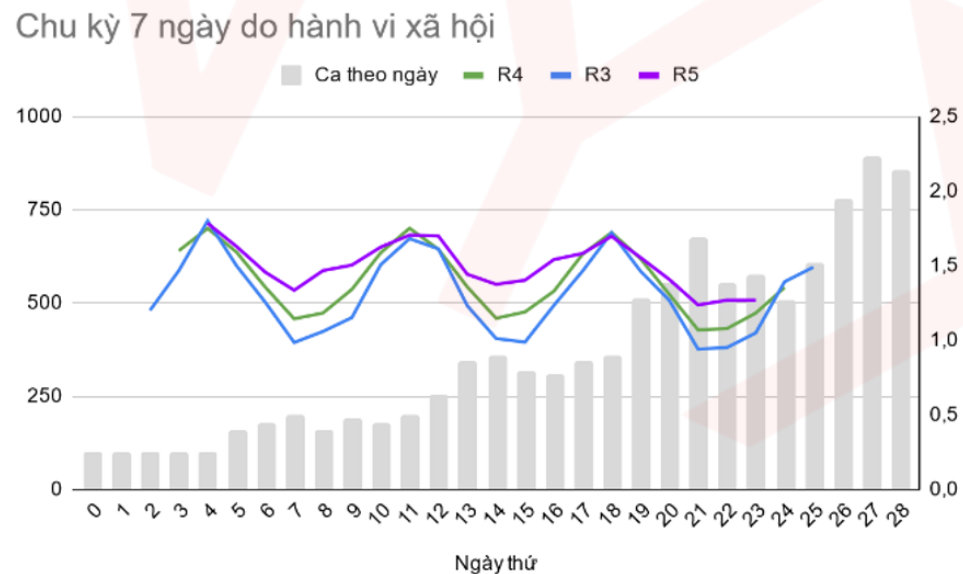
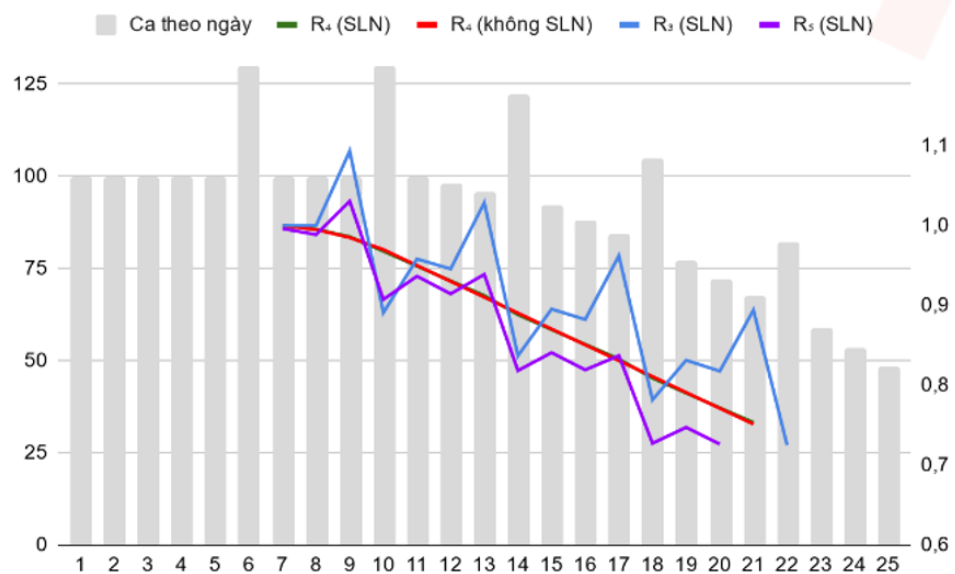


Siêu lây nhiễm

- Chúng tôi đưa ra vài phương pháp để ước lượng số ca bị tăng lên do SLN.
- Phương pháp 1 (cận trên): Tính phần trăm đóng góp bởi SKSLN trong 4 ngày qua (12,9%). Dựa theo mô hình SEIR, thì SLN đóng góp ít hơn nhưng rất gần 12.9% số ca tương lai.
- Phương pháp 2 (cận trên): Cứ mỗi 4 ngày,
$$V(t + 4) = V(t)R_T(t).$$
- Phương pháp 3 (cận dưới): Giống phương pháp 2 nhưng ta chọn giá trị nhỏ nhất của R_T :
$$V(t + 4) = V(t)R_T(t + 3).$$
- Phương pháp 4 (xấp xỉ): Chúng tôi dàn đều số ca trong 4 ngày 11 – 14/07 và 15 – 18/07, sau đó tính theo ánh xạ 1-1 giữa V và R với hoặc không với 1.618 ca theo đường spline.
- Tổng hợp các phương pháp trên, chúng tôi kết luận là SKSLN hôm 08/07 đã tạo ra hơn 10.000 ca cho tới 06/08 ở TP HCM (chưa hiệu chỉnh độ trễ 9 – 10 ngày giữa ngày mắc bệnh và ngày phát hiện). Nó sẽ tiếp tục tạo thêm nhiều ca nữa cho đến khi hết đợt dịch này.

Siêu lây nhiễm

- Chúng tôi cũng mô phỏng 2 ví dụ nhỏ để tìm hiểu hiện tượng.
- SLN khiến cho số ca theo ngày có sóng dương. Đỉnh sóng cách nhau 4 ngày.
- Bài toán SLN tương đương với bài toán xâm nhập. Đường R_T và đường spline trùng nhau. Đây cũng là lý do chúng tôi áp dụng phương pháp thứ 4 trong mục trên.
- Nếu xử dụng sai quãng thời gian (3 hoặc 5 thay vì 4 ngày) để tính R_T thì quỹ đạo R_T lượn sóng và under/over-estimate.
- Hình bên phải cho thấy chu kỳ hành vi xã hội 7 ngày có thể lẫn át chu kỳ liên quan đến khoảng thời gian nối tiếp 4 ngày khi quan sát quỹ đạo R_T .



Thang đo – literature review

- Thế giới hay dùng các chỉ số định lượng sau để đánh giá tình hình:
 - Chỉ số Rt,
 - số ca nhiễm/100 nghìn dân/ngày,
 - Tổng số ca nhiễm.
- QĐ2686 chia vùng thành 4 mức: Nguy cơ rất cao – nguy cơ cao – nguy cơ – bình thường mới. Mỗi mức được tính cho tỉnh/huyện/xã. Thuật toán khá nhiều chi tiết, nên chúng tôi chỉ tập chung chi tiết chính. Thứ nhất, thuật toán được thực tính từ dưới lên: xã, huyện, rồi tỉnh. Thứ hai, thuật toán cho cấp xã xấp xỉ như sau:
 - Bình thường mới là không có ca nào trong 14 ngày.
 - Nguy cơ là có ca F0.
 - Nguy cơ cao là có F0 mà không rõ nguồn lây.
 - Nguy cơ rất cao là có ổ dịch.
- Ưu điểm: an toàn, đi theo tư duy truy vết, phù hợp cho ổ dịch bé và phong tỏa nhanh chóng.

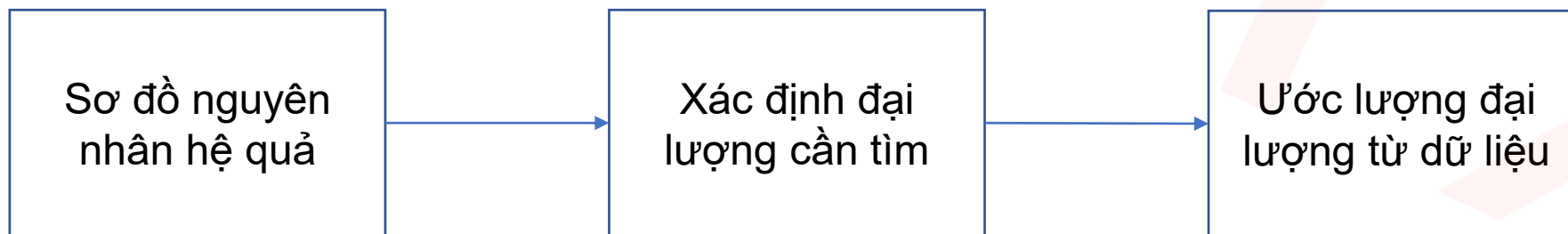
QĐ2686

QĐ2686 có các nhược điểm sau:

- Truy vết là công cụ, còn bệnh dịch bị gây ra bởi vi-rút và tương tác giữa vi-rút và xã hội. Vi-rút và xã hội mới nên là input chính như cách làm trên thế giới. Nó thiếu thông tin về các chủng vi-rút khác nhau và các tỉnh khác nhau nên nó khó thích ứng với chủng mới.
- Gần như đại lượng định lượng duy nhất là đếm ổ dịch. Điều này không phản ánh xu hướng tốt hay xấu vì thiếu thông tin: kích thước ổ và tốc độ tăng giảm kích thước ổ. Một số từ ngữ định tính gây khó đo đếm như: ổ dịch to, nguồn lây khó kiểm soát, bất thường, có nguy cơ lan nhanh, và nằm rải rác. Nó tính từ dưới lên và định tính nên khó áp dụng.
- Khi dịch đã bùng to thì gần như mọi nơi đều thành nguy cơ rất cao. Lúc này ta không phân biệt được nên tập trung vào đâu nữa như tình hình nhiều nơi ở miền Nam bây giờ.
- Ở cấp tỉnh có dùng định lượng là nếu xảy ra “trong 14 ngày, số F0 vượt tỷ lệ 1/100 nghìn dân” thì mức nguy cơ trở lên. Ta xét một ví dụ về một tỉnh có 1 triệu dân, thì 10 ca là đạt tỷ lệ 1/100 nghìn. Như vậy chỉ cần quá 10 ca/2 tuần là thành nguy cơ. Vậy nên chỉ số định lượng trên không có nhiều ý nghĩa.
- Thuật toán phải do trung ương tính toán, địa phương chỉ cung cấp dữ liệu như Chính phủ Đức và nhiều nước làm. Giao nhiệm vụ tính toán cho địa phương với một thuật toán yếu định lượng, sẽ gây khó khăn và mất tính liên thông giữa các vùng.

Một số kiến nghị

- Nghiên cứu trước các chủng vi-rút nếu có thể, ít nhất về R_0 và T . Như Chính phủ Anh có [báo cáo](#) mỗi 2 tuần.
- Cập nhật QĐ2686 với các chỉ số định lượng để phù hợp tình hình mới.
- Đợt dịch này còn dài và còn nhiều lĩnh vực cần tìm hiểu rõ. Cần có một diễn đàn để các chuyên gia (bao gồm cả Việt kiều, chuyên gia độc lập) có thể đóng góp ý kiến và phản biện. Trong sơ đồ dưới đây, chỉ có chuyên gia các ngành [mới làm được bước 1](#). Ngoài ra việc thiết kế một hệ thống thu thập dữ liệu tốt rất quan trọng. [AI và big data](#) rất quan trọng nhưng ở bước 3.



Trân trọng cảm ơn!



Để biết thêm chi tiết, [Website Onyx](#) có thêm 2 slide và nhiều bài viết về COVID-19 (3 bài chính). Nghiên cứu của chúng tôi được báo cáo lần đầu tiên tại Viện Toán học. Cảm ơn sự đóng góp của:

- Nhóm bạn A0 96-99:
 - Trần Tuấn Anh – chuyên gia Giải pháp và Công nghệ
 - Nguyễn Thị Vân Anh – chuyên gia Chính sách công
 - Nguyễn Mai Nguyệt – chuyên gia Quản trị rủi ro
- PGS. TS. Võ Trí Hào – Hiệu trưởng Đại học Gia Định