

Đánh giá diễn biến dịch  
qua số ca mới mắc mỗi ngày

# Đặt vấn đề

- Ví như thực tế xảy ra như sau:
  - Dữ liệu đầu vào chỉ là số ca bệnh mới mắc mỗi ngày và thời điểm nhập viện/cách ly/khỏi bệnh/chết
  - Có các biện pháp can thiệp như giãn cách, khẩu trang và/hoặc cách ly tại các thời điểm khác nhau
  - Có các ca xâm nhập
- Câu hỏi đặt ra là làm thế nào để đánh giá diễn biến dịch?
- 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 (tiếng Anh được tô xanh).
  - Bám sát câu hỏi để tìm loại **reproduction number** phù hợp
  - Những vấn đề khi áp dụng
  - Sơ lược mô phỏng

# Ví dụ 1: trực quan

- **Trường hợp 1:** Tập vật chủ chỉ có 1 cá thể bệnh. Cá thể bệnh này gặp **toàn** cá thể lành không có miễn dịch và sẽ gây ra 10 cá thể bệnh mới (gọi hệ số lây nhiễm cơ bản là  $R_0 = 10$ ).
- **Trường hợp 2:** Giống trên nhưng giả sử 60% tập vật chủ **có miễn dịch** (không thể bị nhiễm bệnh). Cá thể bệnh trên vẫn gặp 10. Về trung bình, 6 cá thể (60% của 10) có miễn dịch nên không bị bệnh, chỉ 4 cá thể còn lại bị bệnh (gọi **effective reproduction number** là  $R_e = R_0 * (1 - 60\%) = 10 * 40\% = 4$ ). Vẫn với  $R_0 = 10$  như trên, nhưng số cá thể bệnh mới là  $R_e = 4$ .  $R_0$  có thể không đổi nhưng  $R_e$  thay đổi khi số cá thể có miễn dịch thay đổi.
- **Trường hợp 3:** Giống trên nhưng giả sử 92% số cá thể có miễn dịch. Về trung bình, 1 cá thể bệnh gây ra  $R_e = 10 * 8\% = 0.8$  cá thể bệnh mới (tương đương với 10 cá thể bệnh gây ra 8 cá thể bệnh mới). Số cá thể bệnh sẽ giảm dần rồi hết. Dịch bị kiểm chế.

# Nhìn nhận vấn đề

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

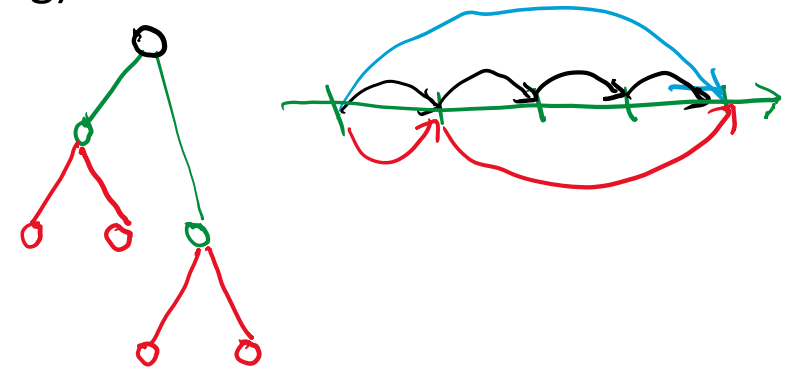
- Môi trường,
  - Đối tượng,
  - Quá trình tăng/giảm số lượng của đối tượng.
- Môi trường là nơi đối tượng có thể tồn tại (người, dơi, .v.v.). Nó thường **hữu hạn**. Ở đây giới hạn trong loài người.
  - Đối tượng có thể gồm nhiều loại loài/hạt/.v.v. cùng một lúc như trong hệ sinh thái. Ở đây là virus.
  - Trường hợp của ta, quá trình tăng/giảm virus được đơn giản hóa thành quá trình tăng/giảm số cá thể bệnh. Tuy nhiên ta vẫn mô hình hóa nồng độ virus trong cá thể bệnh (**virus load**).
  - Quá trình tăng giảm số ca bệnh được tính qua các hệ số lây nhiễm (**reproduction number**). Hệ số lây nhiễm rất trực quan. Nó nói lên là một ca bệnh trung bình sinh ra bao nhiêu ca bệnh mới. Nếu sinh ra ít hơn 1 thì dịch được kiểm soát, nhiều hơn 1 thì dịch lan rộng.

# Reproduction number và threshold value

- **Basic reproduction number**  $R_0$ : the **average** number of secondary infections caused by a **typical** primary infection in an **infinite** and **entirely** susceptible population during its entire period of infectiousness [11].

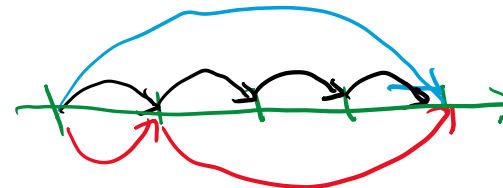
(Các tên gọi khác: basic/case, reproduction/reproductive, number/ratio/rate.)

- **Effective reproduction number**  $R_e(t)$ : the average number of secondary infections caused by a typical primary infection at time  $t$
- **Threshold value**: **threshold**  $< 1$  thì dịch bị kiểm chế và **threshold**  $> 1$  thì dịch lan rộng. Đây là đại lượng ta muốn tìm nếu nó tồn tại.  $R_e$  là một ứng cử viên cho **threshold**.
- Ta chỉ quan tâm  $R_e$  ở đây. Đặc điểm chung của  $R_0$  (o là kí hiệu chung):
  - 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  $R_0$  (**đồng nhất**). Khi đó để tính  $R_0$ , 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  $R_0(t)$ , thì ta có  $n$  biến trong  $n$  ngày. Dữ liệu đầu vào chỉ có  $n$  số cho  $n$  ngày. Do đó  $R_0(t)$  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

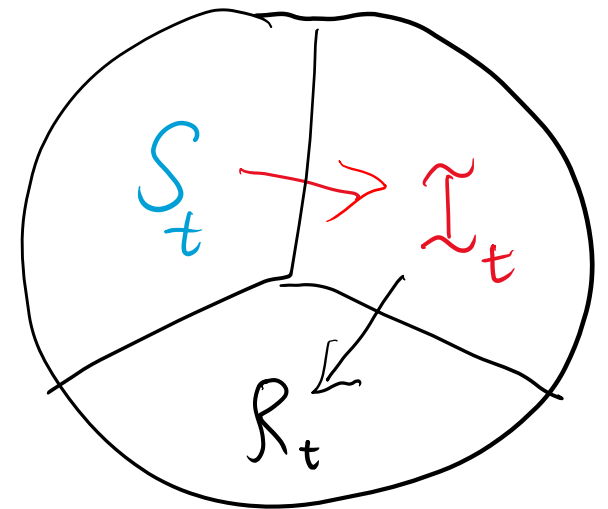
- (Rời rạc) Do dữ liệu là số ca mới mắc mỗi ngày, giả thiết thời gian là ngày ( $t \in \mathbb{N}$ ). Điều này giúp đơn giản hóa mô hình.
- (Độc lập) Để giải quyết vấn đề có nhiều thể hệ tại một thời điểm, giả thiết quá trình truyền bệnh của mỗi cá thể là độc lập. Như vậy quá trình truyền bệnh của một 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 xài để trả lời câu hỏi, chứ không biết ai lây cho ai)



# SIR (thời gian rời rạc)

Nhắc lại mô hình SIR [6]:

- **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  $\mathcal{V}_t$ .

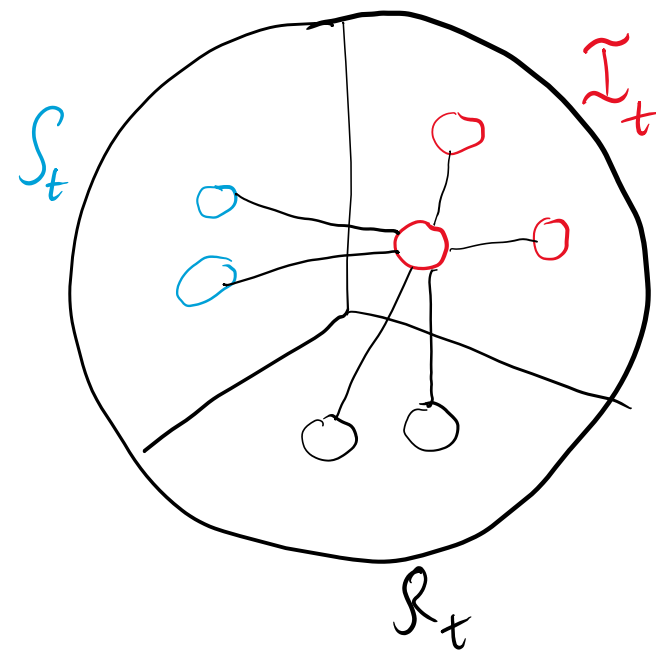
$V_{t+1} = (\text{số cá thể } \mathcal{J}_t) * (\text{số cá thể } \mathcal{S}_t \text{ mà 1 cá thể } \mathcal{J}_t \text{ gặp}) * (\text{tỷ lệ lây})$

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

Ta có

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

- $\beta S_t$  là tốc độ truyền bệnh
- 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.



$$6 \cdot \frac{1}{3} = 2$$

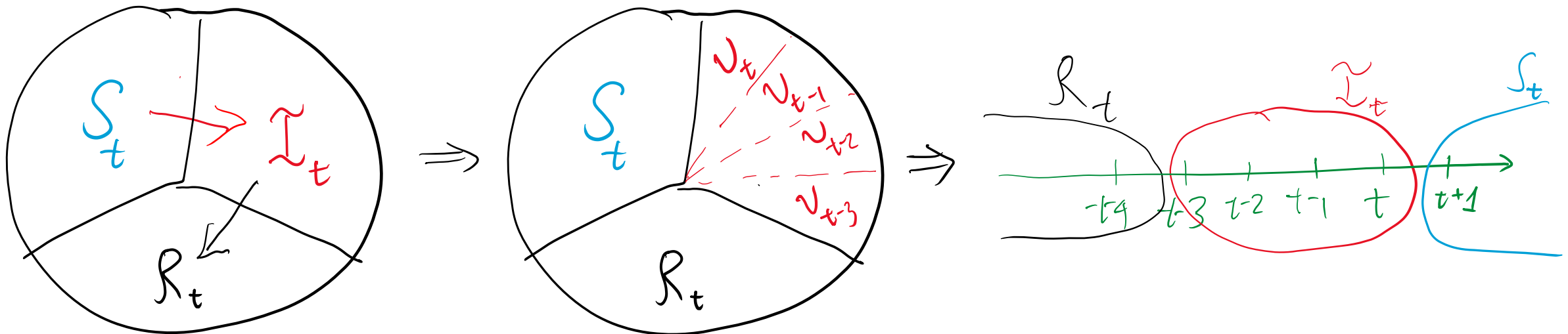
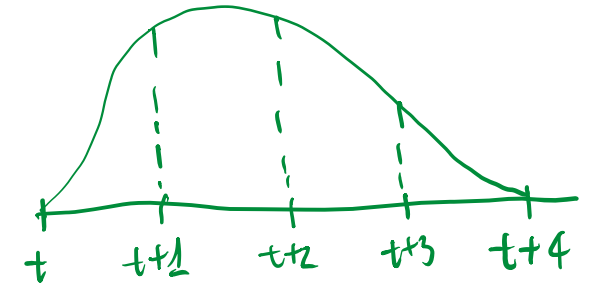
$\nearrow m$        $\nwarrow \frac{S}{N}$



# 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 [3,6]. **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  $I_t = V_t + \dots + V_{t-3}$ .

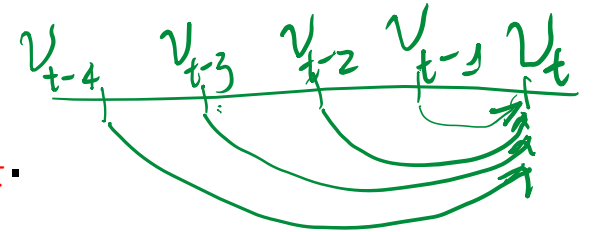


- Từ nay về sau chỉ xét  $t \geq T$ . (Chú ý ta thay  $t + 1$  bởi  $t$  so với trang trước).
- Tương tự SIR, tại thời điểm  $t$  ta có  $\beta_k = mp_k/N$  và  $\mathcal{V}_{t-k}$  lây cho  $(\beta_k S_{t-1})V_{t-k}$  cá thể tại thời điểm  $t$  ( $k = 1, \dots, 4$ ). Do đó

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

- 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  $\beta$  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  $\sum w_k = 1$ .

- Nếu biết  $w$ , từ dữ liệu  $V$  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  $\mathcal{V}_t$ ).

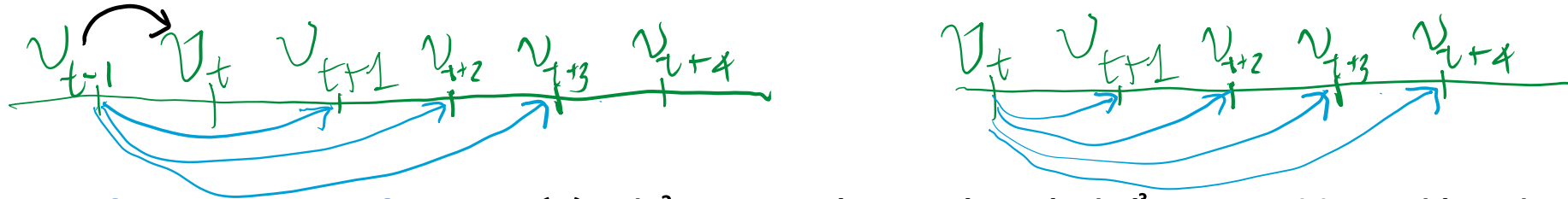
# Tóm tắt mô hình **time-since-infection**

- Với mô hình **time-since-infection**, ta có
  - Một cá thể trong  $\mathcal{V}_s$  lây cho  $R_t w_{t-s}$  cá thể trong  $\mathcal{V}_t$  với  $\sum w_k = 1$  ( $t - T \leq s < t$ )
  - $V_t = R_t (w_1 V_{t-1} + \dots + w_4 V_{t-4}) = R_t \Lambda_t$
  - $\beta_k S_{t-1} = w_k R_t$
- Sau này ta sẽ xét khi nào  $w$  có thể là hằng số
- Bước tiếp theo là định nghĩa và tính các **đại lượng** có thể đánh giá diễn biến dịch. Lẽ tự nhiên là ta xem  $R$  (**instantaneous reproduction number**) có là ứng cử viên, và ta có thể định nghĩa  $R_e$  cho mô hình này không?
- $R_t = \sum R_t w_k$  nên  $R$  là tổng của  $T$  tốc độ lây từ  $t - 4 \rightarrow t$ ,  $t - 3 \rightarrow t$ ,  $t - 2 \rightarrow t$ , và  $t - 1 \rightarrow t$ .  $R$  là số ca bệnh hôm nay nếu mỗi ngày trước có cùng 1 ca bệnh. Điều này không đánh giá được dịch. Ta sẽ bàn thêm sau.
- Về mặt tính toán,  $R$  dễ tính toán. Nó được dùng để dựng quá trình truyền bệnh và tính các đại lượng khác. Nó cũng được dùng để **xấp xỉ hoặc làm cận trên/dưới các đại lượng khác có ý nghĩa thực tế trong các trường hợp đặc biệt**. Ta sẽ tìm các đại lượng đánh giá dịch rồi so sánh chúng với  $R$ .



# Case reproduction number $R_c$

Để đánh giá dịch, ta xem  $\mathcal{I}_t$  sẽ truyền bệnh thể nào sau  $t$ . Có 2 khó khăn. Một là mỗi ca trong  $\mathcal{I}_t$  sẽ sinh ra số ca khác nhau sau  $t$  (một lần lây) tùy theo nó trong  $\mathcal{V}_{t-3}, \dots, \mathcal{V}_t$ . Hai là  $\mathcal{V}_r$  đã sinh ra một số ca trong  $\mathcal{V}_s$  ( $t - 3 \leq r < s \leq t$ ). Sự không đồng nhất này khiến khó tính  $R_e$ . Sử dụng giả thiết độc lập, ta chia bài toán thành  $T = 4$  bài toán nhỏ hơn bằng cách xét 4 nhánh lây nhiễm riêng rẽ.

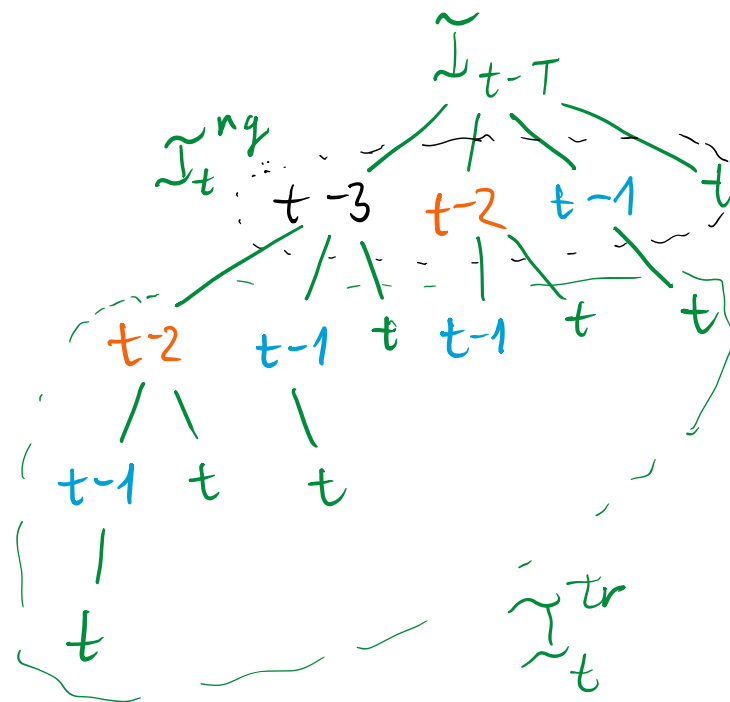
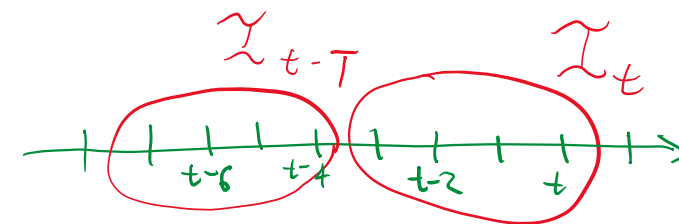


**Case reproduction number**  $R_c(t)$  chỉ quan tâm một cá thể trong  $\mathcal{V}_t$  sẽ lây cho bao nhiêu cá thể qua một lần lây. Một cá thể trong  $\mathcal{V}_t$  được coi là cá thể bệnh số 0 (thế hệ 0). Ta tính số cá thể bệnh được tạo ra bởi nó (thế hệ 1). Ta dùng  $R$  để dựng lại quá trình (một cá thể trong  $\mathcal{V}_t$  lây cho  $R_{t+k}w_k$  cá thể trong  $\mathcal{V}_{t+k}$ ). Ta có

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

Hình dung rằng nếu  $R_c(s) < 1$  trong  $T = 4$  ngày liên tiếp ( $t - 3 \leq s \leq t$ ) thì dịch được kiểm soát tại thời điểm  $t$ . Trang sau sẽ trình bày cách dựng lại quá trình dịch.

- Dựng cây truyền bệnh cho thời điểm  $t$ . Tập các cá thể bệnh là  $J_t$ .
  - Bước 1: Dùng  $R$  và  $w$  để tính thế hệ 0. Mỗi cạnh là 1 lần lây. Xác định các ca trong  $J_t$  bị lây từ bên ngoài  $J_t$  (kí hiệu là  $J_t^{ng}$ ).  $J_t^{ng}$  bao gồm 4 tập con:
    - Tại  $t - 3$ , chính là  $\mathcal{V}_{t-3}$
    - Tại  $t - 2$ , tính qua các lây nhiễm  $t - 6 \rightarrow t - 2, \dots, t - 4 \rightarrow t - 2$
    - Tại  $t - 1$ , tính qua  $t - 5 \rightarrow t - 1$  và  $t - 4 \rightarrow t - 1$
    - Tại  $t$ , tính qua  $t - 4 \rightarrow t$
  - Bước 2: Tiếp tục dùng  $R$  và  $w$  để tính các thế hệ tiếp theo.
- Khái niệm thế hệ ở đây là quy ước. Ta có thể đếm số ca từng thế hệ rồi tính  $R_e$  nhưng nó phức tạp. Cách dễ hơn là chia quá trình truyền bệnh thành 4 quá trình riêng rẽ. Mỗi quá trình xuất phát từ một tập con của  $J_t^{ng}$  được xác định ở bước 1. Nếu  $R_c(s) < 1$  ( $t - 3 \leq s \leq t$ ) thì dịch được kiểm soát **tại thời điểm  $t$** .
- $J_t$  được chia thành 2 tập rời nhau. Tập  $J_t^{ng}$  gồm các cá thể bị lây từ  $J_{t-T}$  (bên ngoài) qua 1 lần lây. Tập  $J_t^{tr}$  gồm các cá thể còn lại trong  $J_t$ . Các cá thể trong  $J_t^{tr}$  lây cho nhau qua 1 đến  $T - 2 = 2$  lần lây.
- Tiếp theo ta xét 2 ví dụ để thấy mối quan hệ giữa  $R$  và  $R_c$ . Sau đó ta tổng kết các tính chất của chúng.



## Ví dụ 2: đồng nhất tốc độ lây

Đồng nhất ở đây theo nghĩa tỷ lệ lây đồng nhất, hay  $p_k = p$ . Do đó  $\beta_k = \beta$  và  $w = 1/T = 1/4$ . Ta có

$$V_t = S_{t-1}(\beta V_{t-1} + \dots + \beta V_{t-4}) = \beta S_{t-1} I_{t-1},$$

$$R_t = \beta_k S_{t-1} / w_k = T \beta S_{t-1},$$

$$R_c(t) = R_{t+1} w_1 + \dots + R_{t+4} w_4 = (R_{t+1} + \dots + R_{t+4}) / T = \beta (S_t + \dots + S_{t+3}).$$

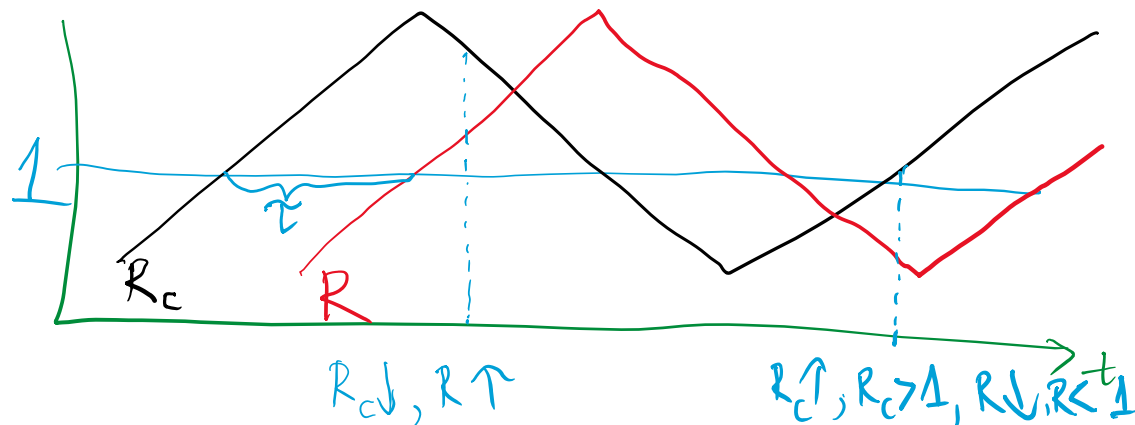
- Trong SIR,  $\beta S_{t-1}$  là tốc độ lây của mỗi cá thể bệnh tại thời điểm  $t - 1$ .  $R_t$  gấp  $T = 4$  lần  $\beta S_{t-1}$  vì  $R_t$  là tổng của 4 tốc độ lây đồng nhất.  $R_t$  giống ước lượng của  $R_e$  trong SIR (so sánh  $T$  với  $1/\gamma$  trong SIR). Tuy nhiên giả thiết về loại bỏ ca bệnh của 2 mô hình khác nhau. Việc loại bỏ ca bệnh giống nhau khi tham số  $\gamma$  trong SIR thay đổi theo thời gian với  $\gamma_t I_t = V_{t-3}$ .
- Nếu  $S$  giảm dần thì  $R_c(t) < T \beta S_t = R_{t+1}$ . Khi các  $S_j$  ( $t \leq j \leq t + 3$ ) xấp xỉ nhau thì  $R_c(t) \approx 4 \beta S_t = R_{t+1}$ .

# Ví dụ 3: một đỉnh

- Cho trước  $0 < \tau = 10 \leq T = 15$ . Giả sử  $w_{10}$  xấp xỉ 1 và các  $w$  khác xấp xỉ 0. Như vậy mỗi ca bệnh mới hầu như chỉ truyền bệnh sau 10 ngày. Ta có

$$R_c(t) = R_{t+1}w_1 + \dots + R_{t+T}w_T \approx R_{t+\tau} = R_{t+10},$$

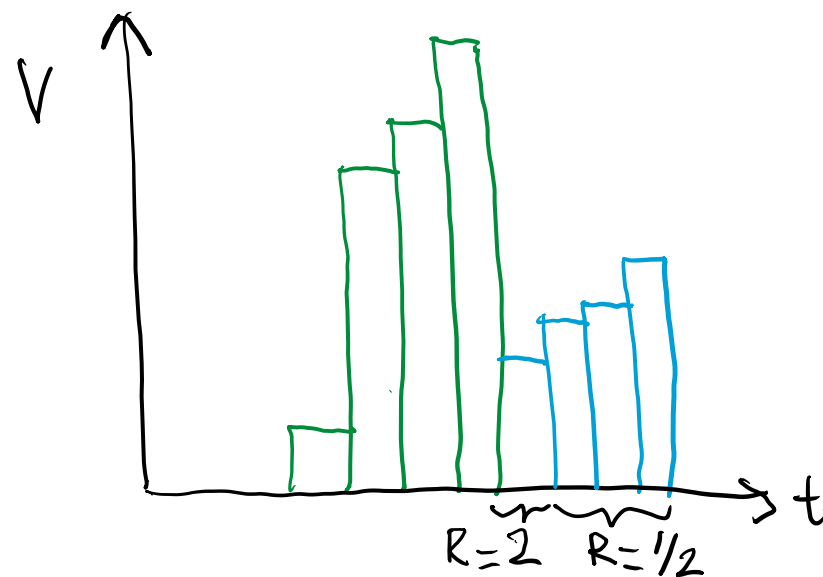
- Như vậy  $R_c$  gần giống  $R$  nhưng trễ hơn  $\tau = 10$  ngày.
- Hình vẽ dưới cho thấy  $R(t)$  tăng/giảm hay to/nhỏ hơn 1 không phản ánh điều tương tự ở  $R_c(t)$ . Mà  $R_c(s)$  ( $t - T + 1 \leq s \leq t$ ) phản ánh dịch bệnh nên  $R(s)$  không phản ánh dịch bệnh.
- Khi  $R$  ít thay đổi trong  $T$  ngày từ  $t + 1$  tới  $t + T$ , thì  $R_c(t) \approx R_{t+1}$  với **mọi**  $w$ . Do đó  $R_{t+1}$  xấp xỉ  $R_c(t)$  trong trường hợp đặc biệt.



- Tiếp theo là mối liên hệ giữa  $R$  và  $V$ . Do  $w_k$  xấp xỉ 0 khi  $k \neq \tau$ , ta có

$$V_t = R_t(w_1V_{t-1} + \dots + w_TV_{t-T}) \approx R_tV_{t-\tau}.$$

- Như vậy trong ví dụ này,  $R_t$  chỉ phản ánh số ca bệnh mới mắc tại  $t$  so với số ca bệnh mới mắc  $\tau$  hôm trước.
- Nếu nói rằng các biện pháp can thiệp thay đổi  $R_t$  là không hoàn toàn đúng. Cho trước  $V_s$  với  $s < t$ , thì  $R_t$  thay đổi khi  $V_t$  thay đổi. Các biện pháp can thiệp tại thời điểm  $t$  có thể không ảnh hưởng mấy  $V_t$ . Ví như giãn cách để giảm gập gở. Trong ví dụ này, nếu chỉ có  $V_{k\tau}$  lớn còn các  $V_s$  ( $s \neq k\tau$ ) rất nhỏ, thì giãn cách ngoài thời điểm  $k\tau$  không ảnh hưởng mấy  $V$ .
- Hình vẽ bên (cho  $T = \tau = 4$  để dễ vẽ,  $V_t = R_tV_{t-4}$ ) cho thấy số ca bệnh giảm so với hôm trước mà  $R > 1$  và ngược lại. Như vậy dùng  $R$  mà chỉ nhìn lại  $\tau - 1 = 3$  thay vì  $\tau = 4$  ngày trước sẽ thấy nghịch lý.





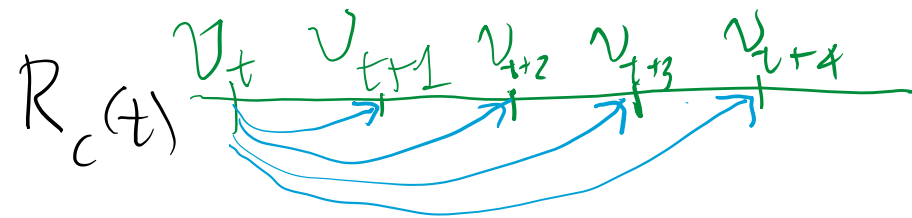
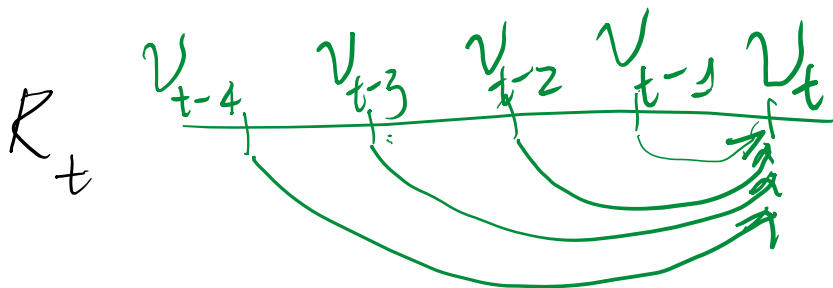
# Tính chất của $R$

Ta thấy vai trò chính của  $R$  là công cụ tính toán. Nó có vai trò thực tế khi nó xấp xỉ các đại lượng khác trong các trường hợp đặc biệt, tuy nhiên vẫn cần điều chỉnh sự khác nhau về thời gian. Nhắc lại các khái niệm liên quan đến hệ số lây nhiễm: thế hệ, tập rời nhau, lây một lần và **threshold**.

- Một cá thể trong  $\mathcal{V}_{t-k}$  lây cho  $R_t w_k$  cá thể trong  $\mathcal{V}_t$  với  $\sum w_k = 1$
- $V_t = R_t (w_1 V_{t-1} + \dots + w_4 V_{t-4}) = R_t \Lambda_t$ . Cho trước  $V$ , để giải được  $R$  thì cần biết  $w$ .
- $\beta_k S_{t-1} = w_k R_t$
- $R$  thỏa mãn các tính chất tập rời nhau và lây một lần, không thỏa mãn các tính chất thế hệ và **threshold**
- $R$  là tổng của  $T$  tốc độ chỉ khi  $w$  đã được chuẩn hóa. Ví dụ  $w$  phụ thuộc cách ly. Khi đó cần chuẩn hóa lại  $w$  và nó khiến  $R$  thay đổi.
- $R_t$  được tính qua  $T + 1$  ngày gần nhất. Do đó chỉ cần  $w$  không đổi trong vòng  $T$  ngày liên tiếp. Muốn tính  $R_t$  trong  $n$  ngày liên tiếp thì cần  $w$  không đổi trong  $T + n - 1$  ngày liên tiếp.

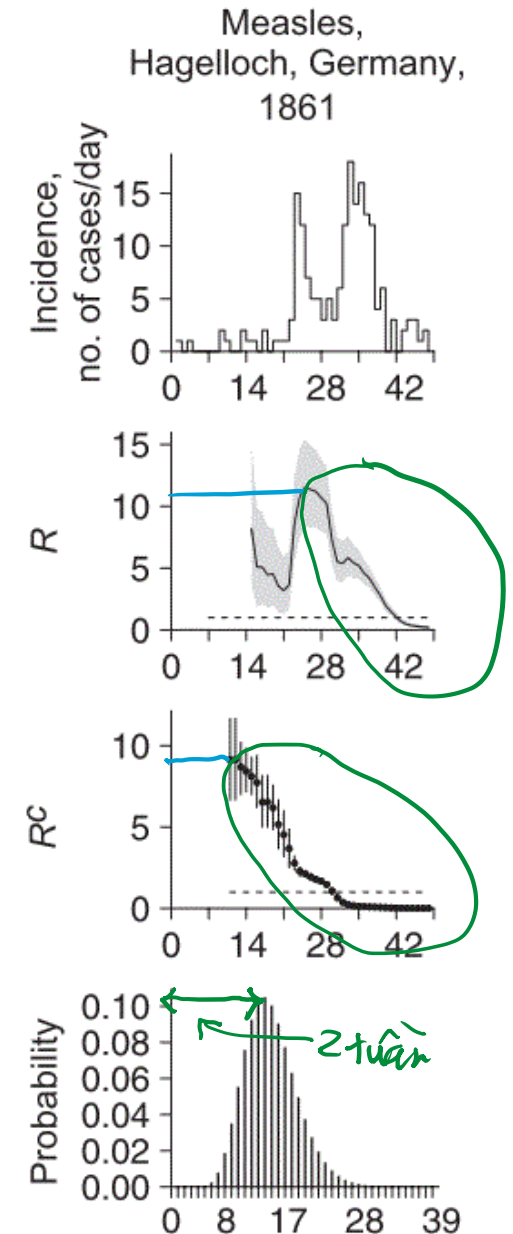
# Tính chất của $R_c$

- $R_c(t)$  chỉ mô tả quá trình xuất phát từ ca bệnh tại  $t$
- $R_c$  có toàn bộ các tính chất thế hệ, lây một lần và **threshold**.  $R_c(t)$  nói về hiện tại ( $t$ ) sinh tương lai qua một lần lây, còn  $R(t)$  lại nói về quá khứ sinh hiện tại ( $t$ ) trong trường hợp số ca mới mỗi ngày quá khứ như nhau. Tuy nhiên do mô hình không giả thiết một quy luật cho  $V$  hay  $R_o$ ,  $R_c$  không có tính dự báo. Ta cần dữ liệu các hôm sau để tính  $R_c$  cho hôm nay.
- $\min\{R_{t+1}, \dots, R_{t+T}\} \leq R_c(t) \leq \max\{R_{t+1}, \dots, R_{t+T}\}$ . Như vậy đỉnh của  $R_c$  thấp hơn đỉnh của  $R$ . Khi  $R_s$  nhỏ (lớn) hơn 1 trong  **$T$  ngày liên tiếp** ( $t + 1 \leq s \leq t + T$ ) thì  $R_c(t)$  nhỏ (lớn) hơn 1.
- Muốn tính  $R_c$  trong  $n$  ngày liên tiếp thì cần  $w$  không đổi trong  $2T + n - 1$  ngày liên tiếp.



# Ví dụ 4: thực tế

- 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  $w$ .
- Giống như trong SIR,  $R_0$  chỉ là một con số phụ thuộc nhiều yếu tố. Cần chuyên gia hoặc các mô hình sâu hơn để biết tác động của từng yếu tố tới  $R_0$ .
- Để tính  $R_c$  thì cần  $w$ , do đó ta xét một hệ số lây nhiễm khác mà chỉ cần  $T$  để tính.

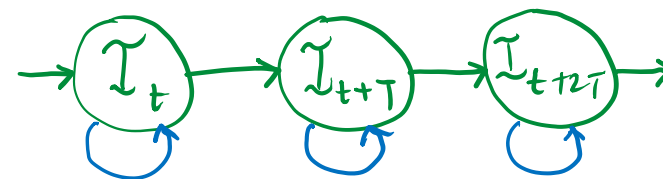
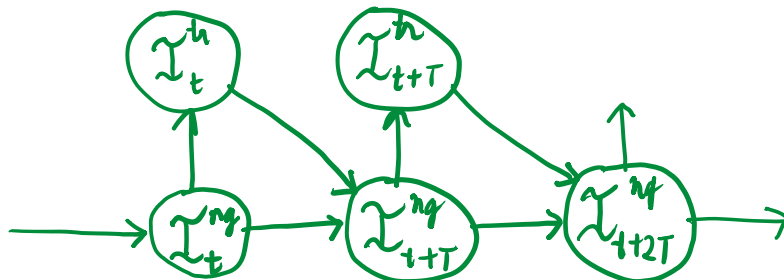


# Reproduction number $R_T$

- Nhắc lại  $\mathcal{J}_t$  tách thành 2 tập  $\mathcal{J}_t^{ng}$  và  $\mathcal{J}_t^{tr}$ . Ta biểu diễn quá trình lây bệnh thành các khối như hình vẽ dưới. Các mũi tên xanh lá cây biểu diễn lây 1 lần. Các cá thể trong  $\mathcal{J}_t^{ng}$  không lây cho nhau. Các cá thể trong  $\mathcal{J}_t^{tr}$  lây cho nhau qua 1 đến  $T - 2$  lần lây. Ta tiếp tục đơn giản hóa bằng cách gộp  $\mathcal{J}_t^{ng}$  và  $\mathcal{J}_t^{tr}$  thành  $\mathcal{J}_t$  (mũi tên xanh da trời biểu thị lây 1 hoặc nhiều lần). Tại mỗi thời điểm  $t$ , ta có một sơ đồ các block khác nhau (các sơ đồ trùng lặp với chu kỳ  $T$ ). Gọi ( $I_t > 0$ )

$$R_T(t) = \frac{I_{t+T}}{I_t} \text{ (dịch đi 4 ngày so với [8])}.$$

- $R_T$  có tính chất tập rời nhau và **threshold**, không có tính chất thể hệ và lây một lần.
- Sơ đồ khối có tính chất sau:
  - Mỗi khối giống một thể hệ. Mỗi khối chỉ có 1 mũi tên đi vào và 1 mũi tên đi ra. Sơ đồ không có chu trình khép kín.
  - Các khối có cùng kích thước và cấu trúc. Ví dụ nếu mỗi ca chỉ sinh ra 1 ca, thì việc cùng kích thước và cấu trúc đảm bảo  $R_T = 1$ .
  - Nếu chọn quãng thời gian dài hơn cho mỗi khối thì kém hiệu quả, còn ngắn hơn thì có thể có nhiều mũi tên đi ra từ 1 khối.



# Tham số thay đổi theo thời gian

**Time-since-infection** cho phép tham số thay đổi [3]. 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.
- Khi  $m_t, p_{k,t}, N_t$  và  $S_t$  thay đổi mà  $p_{k,t}:p_{1,t}$  không đổi, thì  $w$  không đổi. Nếu các sự thay đổi đó ảnh hưởng  $V_t$  (không phải lúc nào cũng ảnh hưởng như trong ví dụ 4) thì  $R_t$  thay đổi và phản ánh sự tác động đó.
- Tính đồng nhất vẫn giữ nguyên do cùng tốc độ truyền bệnh cho các cá thể trong  $\mathcal{V}_t$ .

# Vấn đề khi ước lượng $R$

Vì  $R$  biến đổi khá nhiều, nên [Cori et al. 2013] giả sử  $R$  không đổi trong  $\tau$  ngày

$$R_t = R_{t-1} = \dots = R_{t-\tau+1} =: R_{t,\tau}.$$

Mỗi thời điểm có một mô hình riêng cho  $R$ . Chú ý rằng  $\frac{x}{u} = \frac{y}{v} \implies \frac{x}{u} = \frac{x+y}{u+v}$ . Do đó nếu  $\Lambda > 0$ , từ  $R_t = V_t/\Lambda_t$ , ta có

$$\hat{R}_t = (V_t + \dots + V_{t-\tau+1})/(\Lambda_t + \dots + \Lambda_{t-\tau+1}).$$

Với Bayesian thì công thức hơi khác

$$\hat{R}_t = (a + V_t + \dots + V_{t-\tau+1})/(b^{-1} + \Lambda_t + \dots + \Lambda_{t-\tau+1})$$

với  $a = 1$  và  $b = 5$  (có thể áp dụng cùng suy luận trên với  $\frac{x}{u} = \frac{y}{v} = \frac{a}{1/b}$ ). Họ áp dụng vào các dữ liệu với  $T \geq 10$  nên họ chọn  $\tau = 7$ .

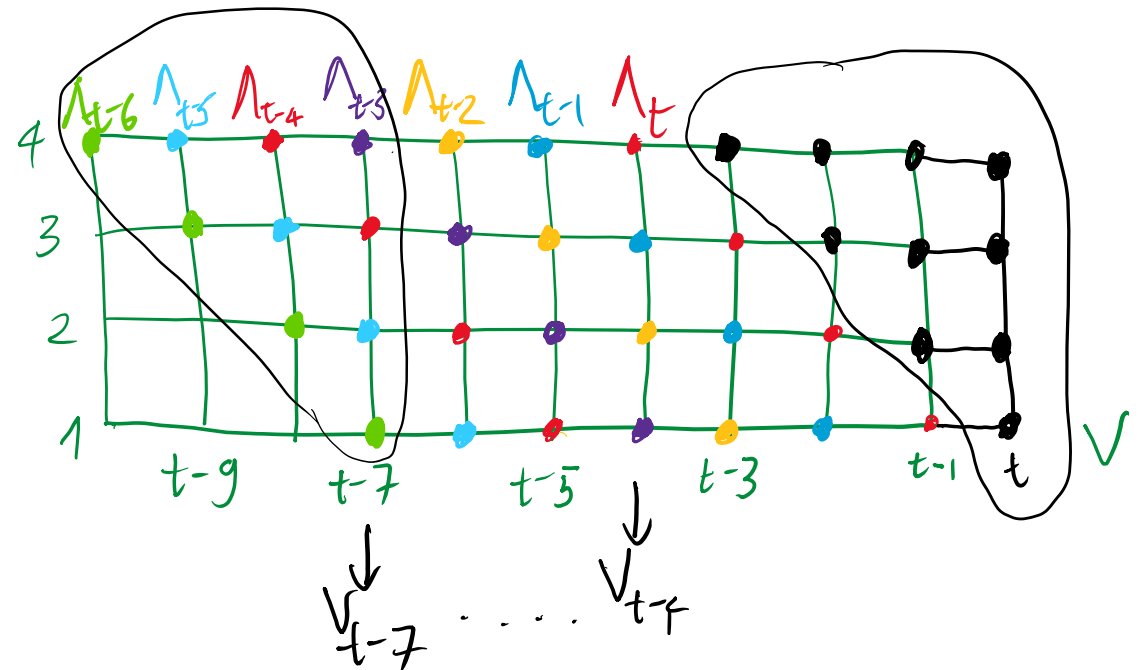
Một vấn đề với  $\hat{R}_t$  là nó là một tỷ số mà tử số và mẫu số có chung một số  $V_s$ . Phần chung càng nhiều khi  $T$  càng nhỏ và/hoặc  $\tau$  càng lớn. Điều này làm mất đi tính tập rời nhau. Trong khi  $R$  so sánh hôm nay với quá khứ  $T$  ngày gần nhất, thì  $\hat{R}_t$  so sánh “hôm nay và quá khứ gần” với “quá khứ xa” qua một phần chung. Hình vẽ tiếp theo sẽ mô tả rõ hơn.

Ta xét ví dụ với  $T = 4$  và  $\tau = 7$ . Không công bằng vì  $T > \tau$  nhưng ta muốn vấn đề dễ nhìn hơn. Nhắc lại  $\Lambda_t = w_1 V_{t-1} + \dots + w_4 V_{t-4}$ . Hình vẽ dưới biểu diễn mỗi khoản  $V_{t-k} w_k$  thành 1 điểm có tọa độ  $(t-k, k)$ . Mẫu số  $\Lambda_t + \dots + \Lambda_{t-\tau+1}$  là tổng các khoản này. Nếu ta có đủ 4 khoản trên một hàng dọc qua  $t-s$  thì tổng của chúng bằng  $V_{t-s}$  do  $\sum w_k = 1$ . Như vậy  $\Lambda_t + \dots + \Lambda_{t-\tau+1}$  bao gồm  $V_{t-7} + \dots + V_{t-4}$ . Tóm lại:

- Phần chung nhau là các chấm không được khoanh vùng. Gọi tổng là  $C_t$ .
- Phần chỉ có ở tử số là vùng khoanh các chấm đen. Gọi tổng là  $A_t$ .
- Phần chỉ có ở mẫu số là vùng khoanh các chấm màu. Gọi tổng là  $B_t$ .
- Khi đó

$$\hat{R}_t = \frac{A_t + C_t}{B_t + C_t} = \frac{1 + A_t/C_t}{1 + B_t/C_t}$$

- Khi phần chung  $C_t$  càng lớn thì  $\hat{R}_t$  càng bị làm phẳng

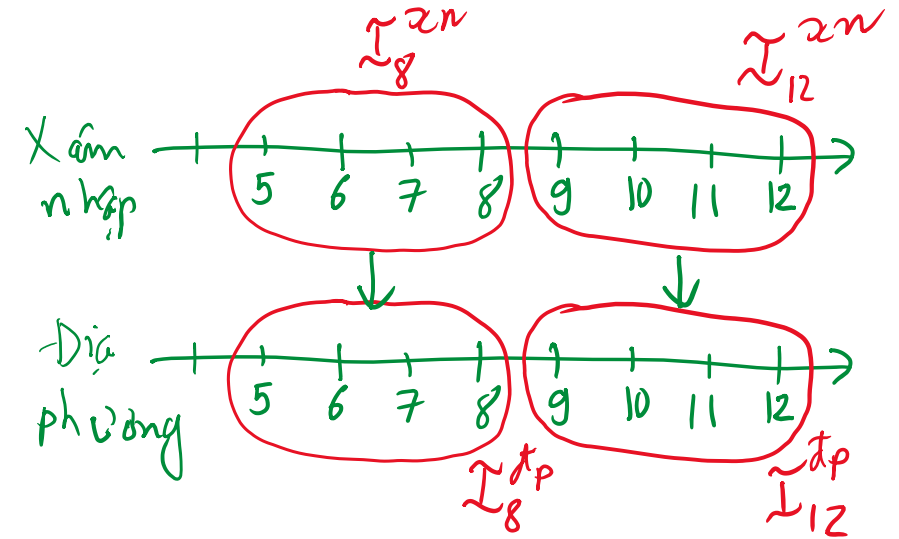


# Xâm nhập

- 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  $J_t^{\text{xn}}$  và  $J_t^{\text{ng}}$ ). Vì  $J_{t=12}^{\text{xn}}$  không được tạo ra bởi  $J_8^{\text{đp}}$  và  $J_8^{\text{xn}}$ , 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. Ta lấy cận trên để ước lượng bằng cách tính cả các 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. Phương pháp này có lẽ có nhiều vấn đề hơn do cách giải quyết cá thể xâm nhập ở phần chung nhau giữa tử số và mẫu 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 [8], 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. Ta không trình bày ngẫu nhiên hóa ở đây vì nó tùy thuộc vào cách ta thu thập dữ liệu.

# Tổng kết lý thuyết cho time-since-infection

- Về các giả thiết:
  - Thời gian là ngày ( $t \in \mathbb{N}$ ). Các sự kiện gặp gỡ cũng như lây bệnh chỉ xảy ra tại các thời điểm cách nhau đúng 1 ngày.
  - Tại thời điểm  $t + 1$ , một cá thể gặp  $m_t > 0$  cá thể bất kì. Quá trình gặp gỡ giữa các cá thể là độc lập.
  - 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  $p_{k,t}$  cho mỗi gặp gỡ sau  $k$  ngày (gặp gỡ tại thời điểm  $t + k$  với  $k = 1, \dots, T$ ), trong đó  $\sum p_{k,t} > 0$  và  $p_{1,t} : \dots : p_{T,t}$  không đổi.
  - Không cá thể bệnh nào bị loại bỏ trong vòng  $T$  ngày kể từ khi mới mắc bệnh.
- Về các hệ số lây nhiễm (tính  $R_0$  trực quan hơn SIR vì sự loại bỏ ca bệnh là đồng nhất)
  - Chỉ cần một sơ đồ khối với các khối tách rời và kích thước các khối
  - $R_c(t)$  là chính xác, nhưng chỉ cho  $\mathcal{V}_t$  và cần biết  $w$ . Cần tính cho  $T$  ngày liên tiếp.
  - $R_e$  khó tính
  - $R_T(t)$  là chính xác, cho toàn bộ  $\mathcal{I}_t$ , không cần  $w$ , nhưng cần biết  $T$
  - $R$  chỉ nên dùng trực tiếp khi nó có thể xấp xỉ  $R_c$  và cần chỉnh thời gian
  - Toàn bộ các hệ số lây nhiễm không có tính dự báo vì phương pháp là **data driven**. Để dự báo cần một quy luật. Ví dụ khi dịch mới bắt đầu và không bị can thiệp, ta hay giả thiết  $V$  tăng theo tốc độ mũ. Quy luật cho  $V$  sẽ dẫn đến quy luật cho  $R_0$ .

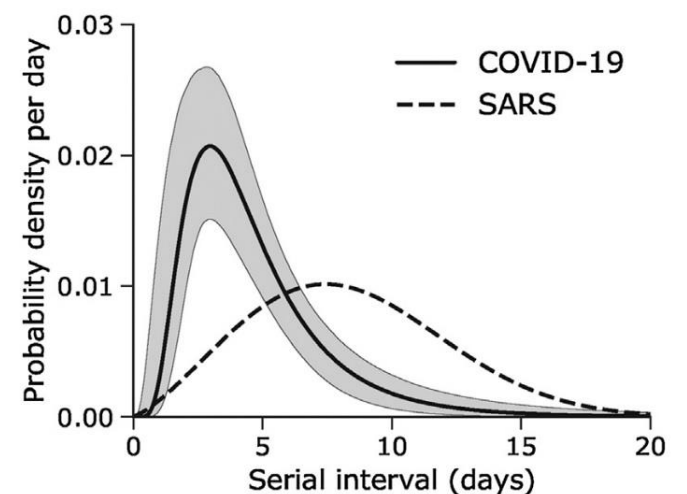
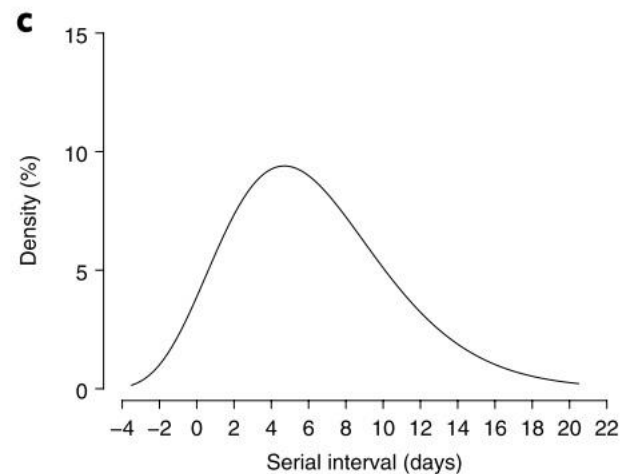
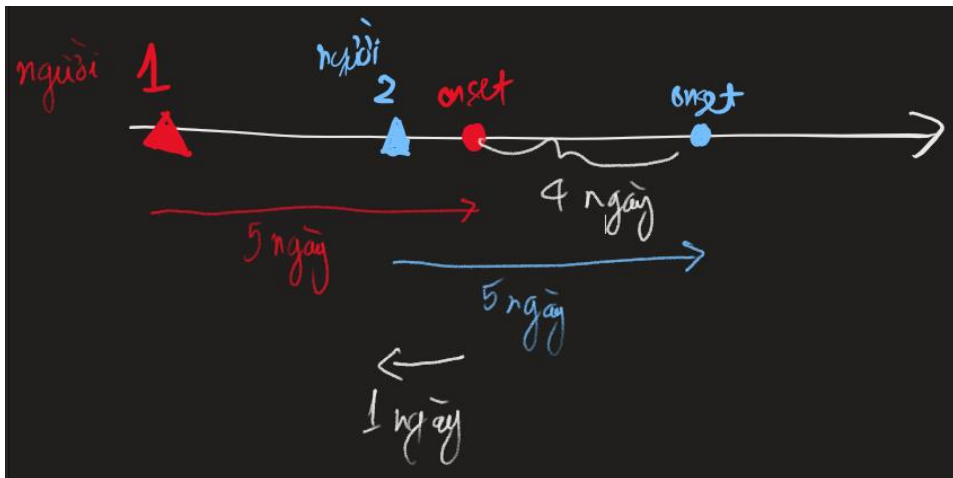
# Ứng dụng

- 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) [4,7]
- 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]).



# Ví dụ 5: Phát hiện ổ

- (Kích thước ổ) Giả sử không phải ai có triệu chứng cũng đi khám nên xác suất nhập viện chỉ là 10%. Giả sử xác suất chẩn đoán ra là 50% khi nhập viện. Khi đó xác suất tìm ra người bệnh là  $p = 5\%$  (5% khá phù hợp vì xác suất chết cũng đã 2-3%). Xác suất để ổ dịch có  $n$  người mà vẫn chưa tìm ra ổ là  $(1 - p)^n$ . Vậy là cần  $n = 60$  thì xác suất này mới là  $0.95^{60} = 4.6\% < 5\%$ . Tức là 95% sẽ phát hiện ra ổ khi có đến 60 ca.
- (Quy đổi số ca ra số ngày) Giả sử ta ước lượng được  $R_0 = 2$  và  $T = 4$ , với  $T$  là trung bình của SI. Giả sử  $p = 5\%$ . Sau  $x * T$  ngày kể từ khi F0 phát bệnh, từ 1 ca sẽ lên thành số ca

$$[R_0^{x+1} - 1] / [R_0 - 1] \approx R_0^{x+1} / (R_0 - 1).$$

Khi  $20 = 5 * 4$  ngày đã trôi qua kể từ khi F0 phát bệnh ( $x = 5$ ), thì từ 1 thành  $2^6 - 1 = 63$  ca. Nếu tính cả 5 ngày ủ bệnh, và 5 ngày để từ phát bệnh thành trở nặng đi viện, thì 95% sẽ tìm ra ổ mới sau  $20 + 5 + 5 = 30$  ngày. Sau 30 ngày ( $30 = 7.5 * 4$ ), thì số ca bệnh đã là  $2^8 - 1 = 255$ .

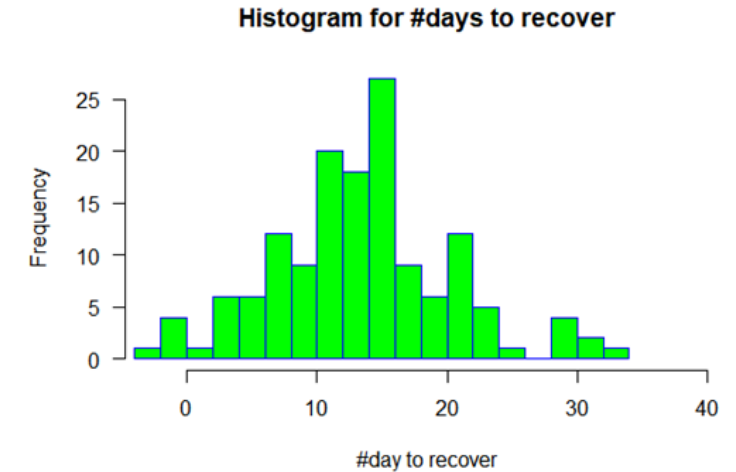
- Chú ý rằng sau 9 ngày thì khả năng truyền bệnh là rất thấp, do đó không nên lo lắng về F0 sau 9 ngày. Cũng không nên lo lắng các ca không triệu chứng khi mới phát hiện ổ. Mục đích là giảm  $R_0$  xuống dưới 1. Ví dụ nếu phát hiện ra một ổ, coi mọi cá thể bệnh lúc mới phát hiện ổ là các F0. Tìm và diệt gần hết toàn bộ chuỗi lây nhiễm của F0 tìm ra (bao gồm cả ca không triệu chứng), chứ không nhất thiết phải dò ra tất cả các F0 (vì có ca không triệu chứng).

# Ví dụ 6: Việt Nam

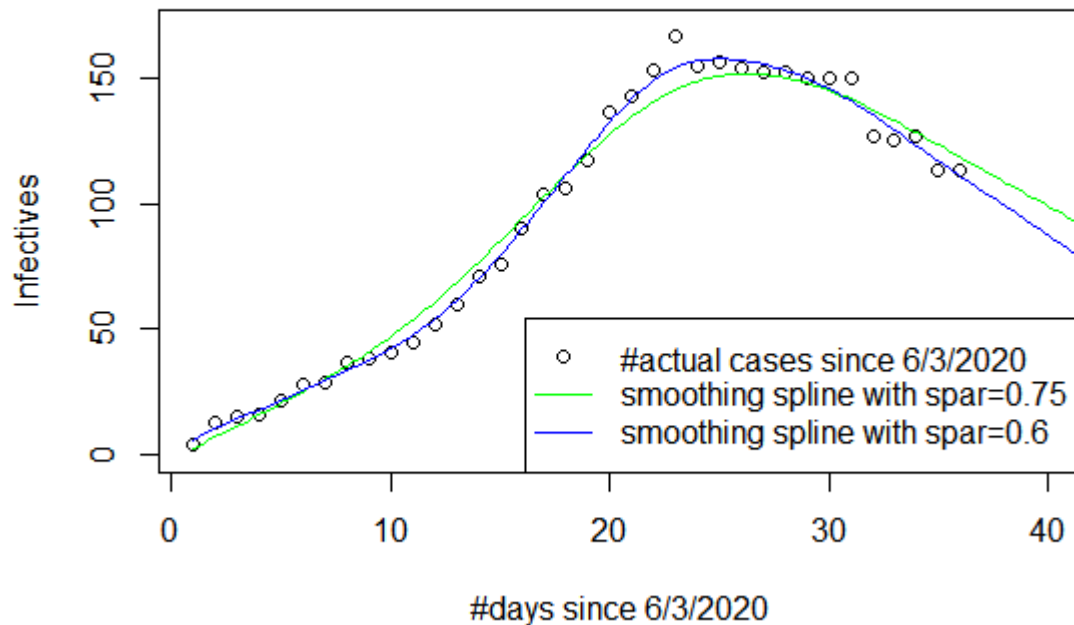
- Dữ liệu
  - Covid-19 trong Việt Nam cho đến ngày 10/04/2020 từ file “Theo doi NCOVI 2020.04.11.xlsx”.
  - Từ 13/02/2020 đến 06/03/2020 mới phát hiện ca bệnh tiếp theo. Khoảng thời gian đủ dài để không có mối liên hệ nên ta xóa mọi ca trước 06/03/2020.
  - Xóa mọi ca phát hiện ngay lúc nhập cảnh, hoặc trong khu cách ly.
  - Khi đó còn **132 ca trong 36 ngày**.
  - Cho các ca **xâm nhập**, dùng số liệu tổng hợp từ file “Nhờ\_Simulation.xlsx”.
- **Giả sử**
  - Dữ liệu thu thập đầy đủ và “Ngày xác nhận” là ngày **onset**.
  - SI dương. Thực tế SI có thể âm.
  - SI trung bình là 4 ngày và có thể đến 20 ngày. Do biện pháp **contact tracing** làm thay đổi quá trình truyền bệnh, ta giả sử  $T = 4$ . Nghĩa là bị cách ly/nhập viện trong vòng 4 ngày sau khi **onset**.

# Số ca bệnh đang điều trị tích lũy theo ngày

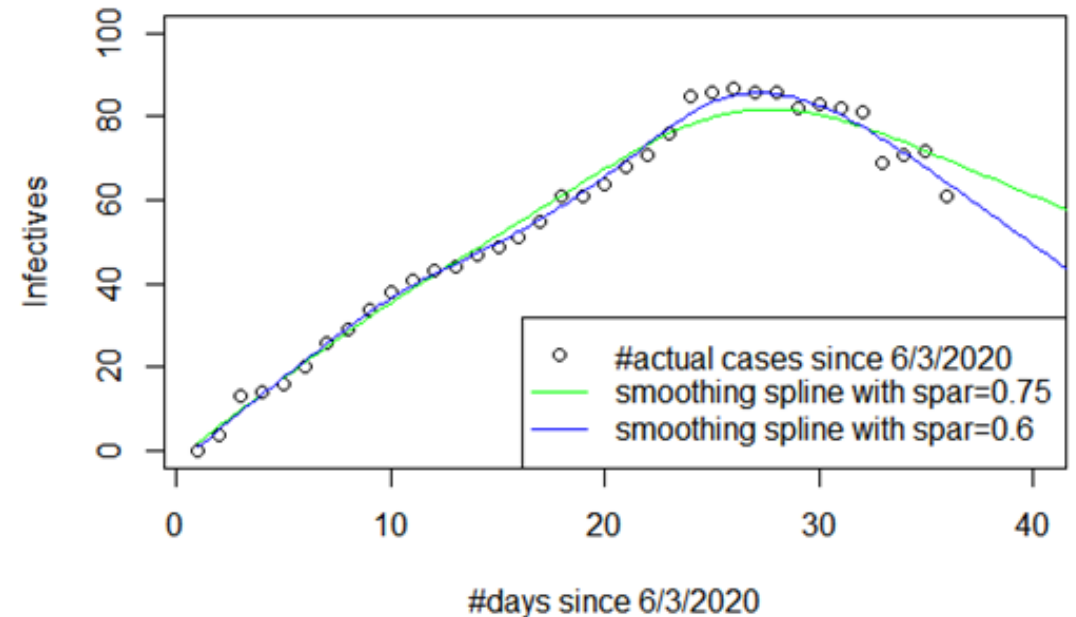
- Hai đồ thị phía dưới biểu thị số ca bệnh **đang** điều trị tích lũy theo ngày. Sau khi xóa cách ly (hình góc dưới bên phải), dáng hai đồ thị vẫn khá giống nhau. Tuy nhiên sau khi xóa, đỉnh bị lùi lại và tốc độ tăng thấp hơn nhiều.
- Nếu dùng dữ liệu này, người bệnh được cho vào tập loại bỏ chỉ khi khỏi bệnh. Điều này không phù hợp vì số ngày điều trị đa phần lớn hơn 4 nhiều (hình góc trên bên phải với một số khoảng âm).



#infectives by day since 6/3/2020

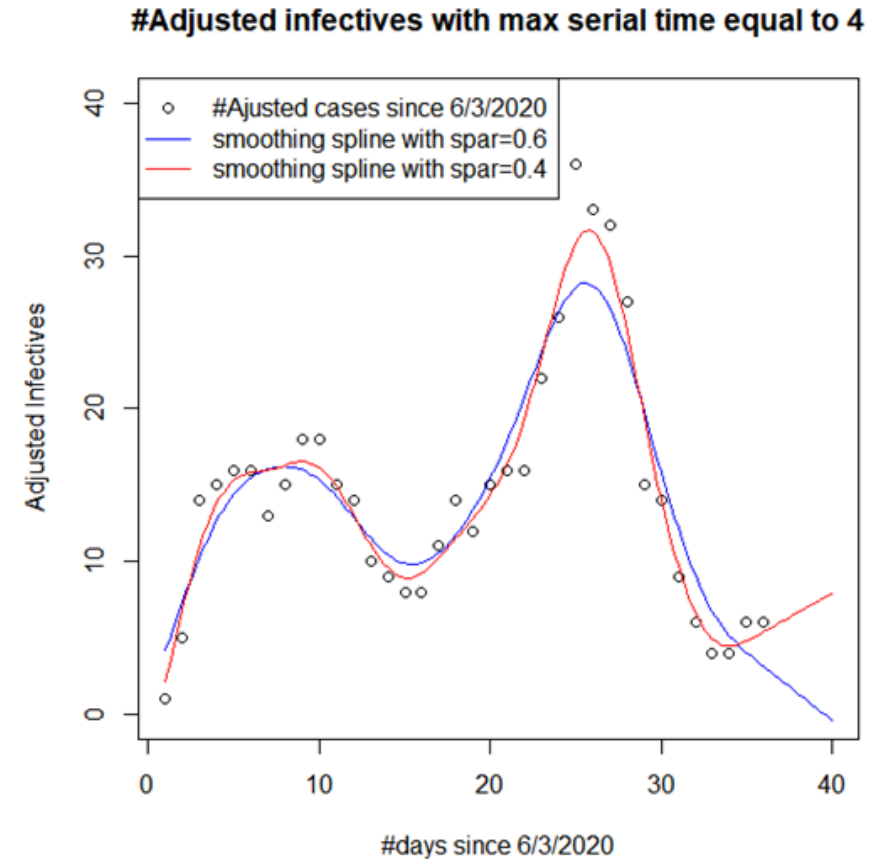
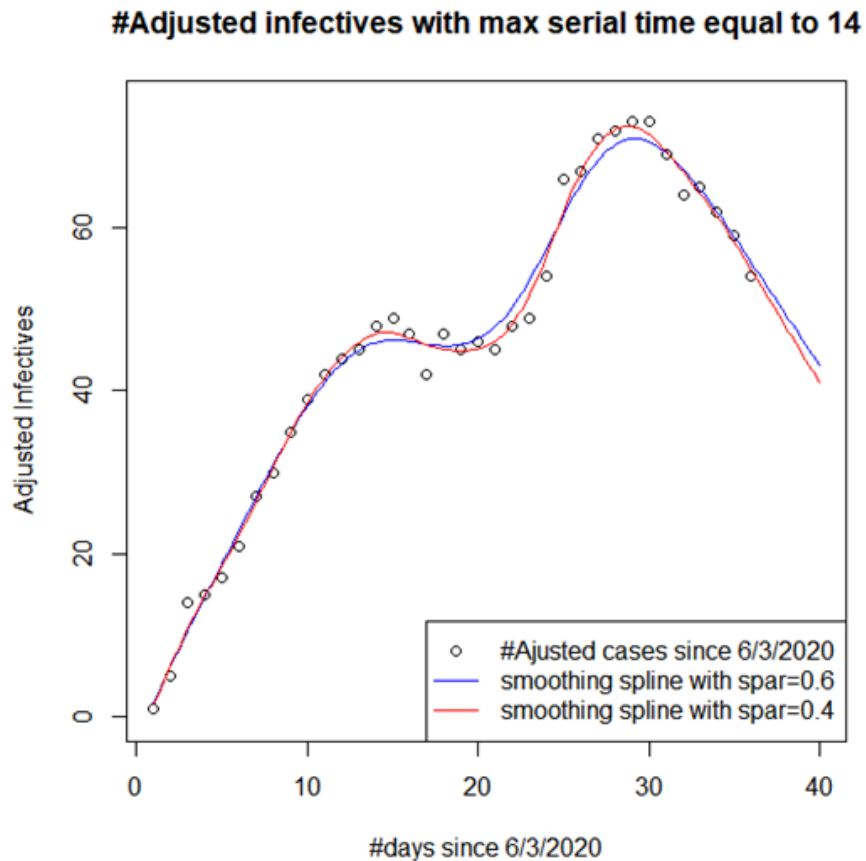


#infectives by day since 6/3/2020 (exclude quarantine)



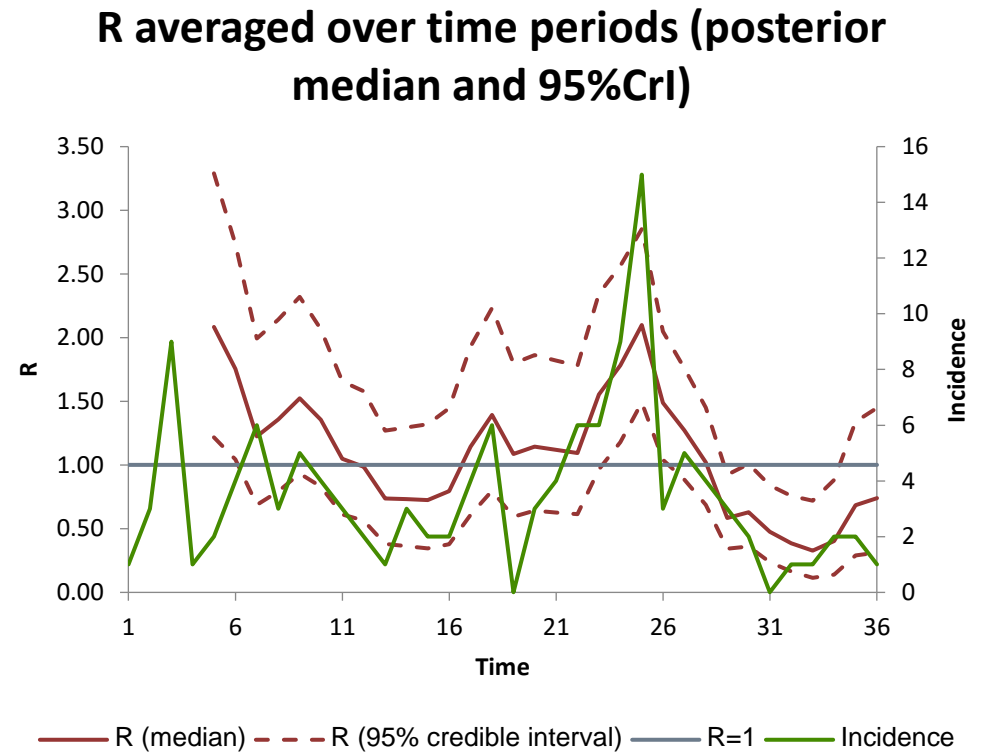
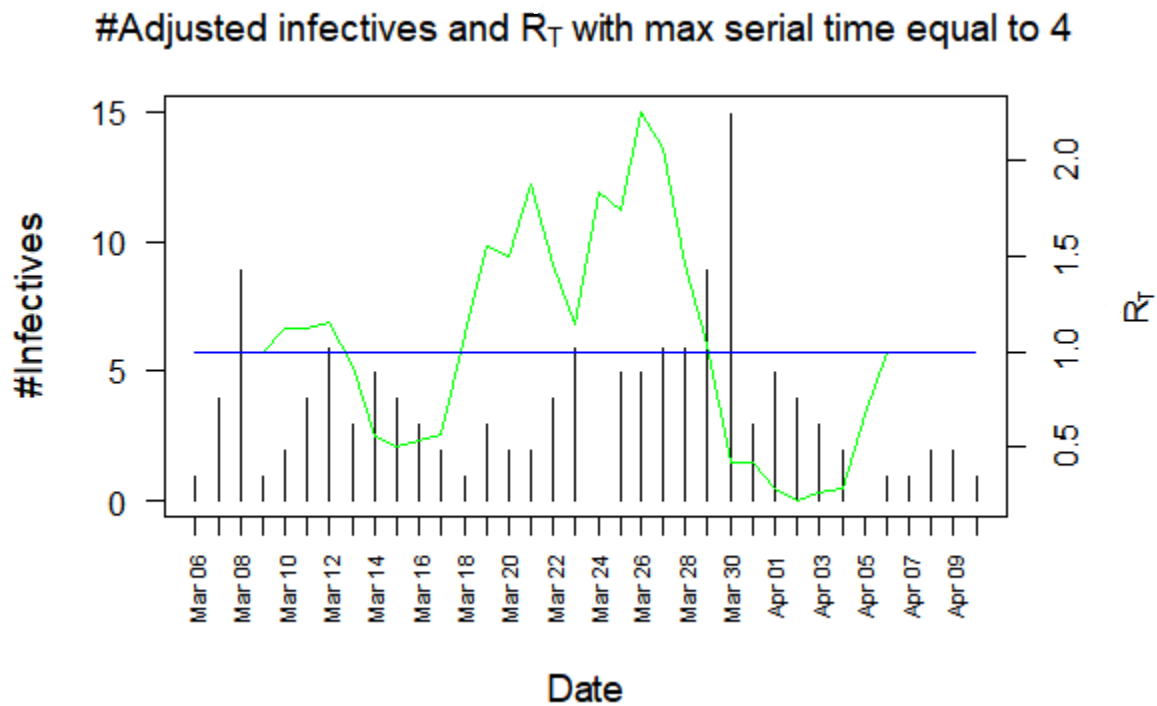
# Loại bỏ ca bệnh trong vòng 4 hoặc 14 ngày

- Số ca bệnh trong tập  $\mathcal{I}_t$  với  $T = 14$  (bên trái) và  $T = 4$  (bên phải). Cả hai đồ thị dường như có hai đỉnh. Nghĩa là dịch được kiểm soát và mang dáng dấp của các cụm dịch.



# $R_T$ và $\hat{R}$

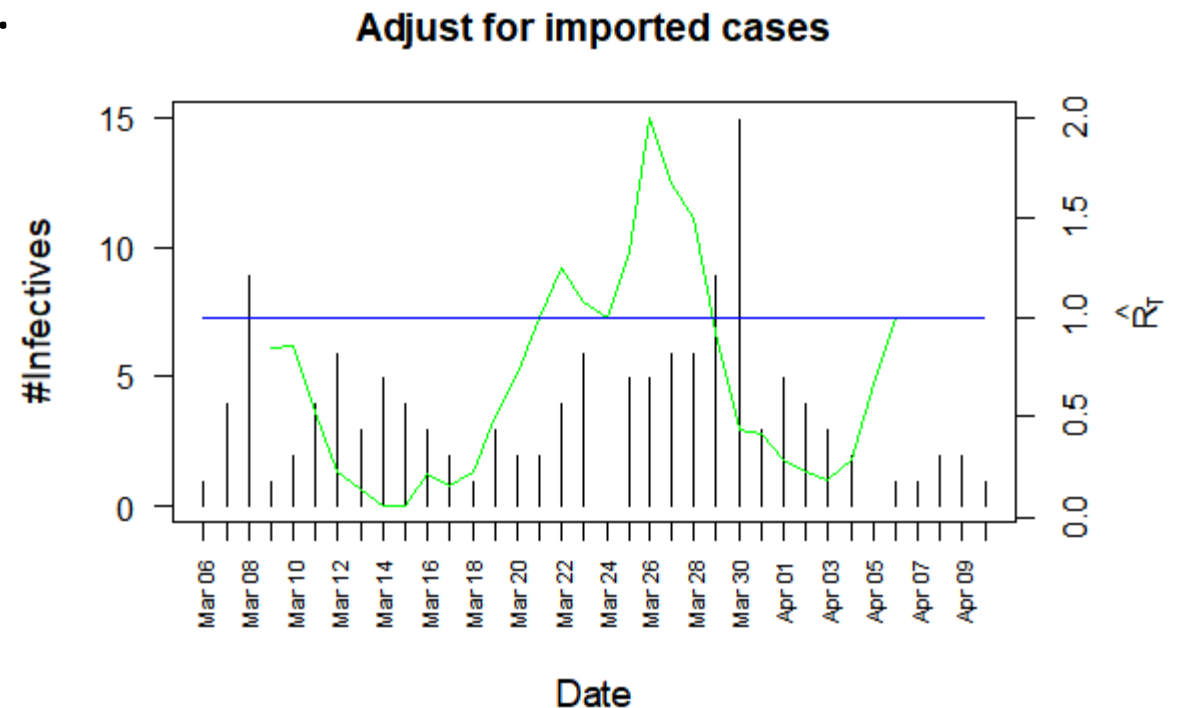
- Chưa xử lý xâm nhập ở đây
- Bên trái là đồ thị cho  $R_T$ . Không có **confidence/credible interval** vì giả thiết dữ liệu đầy đủ.
- Bên phải là phương pháp trong [1] với  $\tau = 4$  và (**mean SI, standard deviation of SI**) = (4,4).





# Xâm nhập

- 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 thì dùng  $\hat{R}_T$  để làm mô phỏng là không phù hợp



# Xem lại một số giả thiết

- SI dương: khi SI âm thì một cá thể trong  $\mathcal{J}_{t+T}$  trong khi nó nên trong  $\mathcal{J}_t$  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].
- Nếu SI dương, và **incubation time** và SI độc lập, thì cần dịch  $\hat{R}_T$  về quá khứ 5 ngày (trung bình **incubation time** là 5 ngày). Như vậy ta ước lượng  $R_T$  chậm hơn hiện tại **9 ngày** (5 ngày **incubation time** và 4 ngày phía trước).
- Nếu thực tế  $T > 4$  thì ta có thể tăng  $T$  lên.

# Mô phỏng với 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  $R_T$ ). Ta lấy ví dụ với SIR.

- Kí hiệu  $t = 0$  là hôm nay. Để simulation với mô hình SIR thì cần 5 tham số đầu vào:  $\beta$  (đã trình bày),  $\gamma > 0$  (tốc độ loại bỏ ca bệnh),  $S_0$  (số người trong  $S_{\text{hôm nay}}$ ),  $I_0$  (số cá thể bệnh hôm nay), và  $N$  (tổng số cá thể).
- Trong SIR, ta có  $R_e = \beta S_0 / \gamma$ . Biết  $R_e$  thì ta chỉ cần xác định 4 tham số. Ta lựa chọn 4 tham số dễ xác định nhất qua các nghiên cứu khác. Thường cố định  $N$ , lấy  $I_0$  từ dữ liệu, lấy  $\gamma$  từ nghiên cứu khác, và cuối cùng là tính  $S_0$ . Khi đó  $\beta$  được tính từ  $R_e = \beta S_0 / \gamma$ .
- Khi dịch bệnh mới bắt đầu, có thể giả thiết  $S_0 \approx N$ . Khi dịch bệnh đã lan rộng, có thể cần các phương pháp như sero-survey để ước lượng  $S_0$ .
- 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 [2].
- Sau khi có thông tin trên thì có thể dùng bất cứ ngôn ngữ lập trình nào có gói giải hệ phương trình vi phân như ngôn ngữ R với package deSolve.

# R code cho SIR [9]

# Chỉ vài dòng code và rất dễ dùng:

```
library(deSolve)
#---parameter-----
b <- 1.4 #beta
g <- 0.3 #gamma
parameters <- c(bet=b,gamm=g)
init <- c(S=1-1e-6, I=1e-6, R=0)
time <- seq(0, 70, by=0.5) #every 0.5 day
#---equation-----
eqn <- function(time, state, parameters){
  with(as.list(c(state, parameters)),{
    dS <- -bet*S*I
    dI <- bet*S*I-gamm*I
    dR <- gamm*I
    list(c(dS,dI,dR))  })}
#---output (deSolve: ode)-----
out <- ode(y=init, times=time, eqn, parms=parameters)
```

# Mô phỏng contact tracing

- [Hellewell *et al.*, 2020] xem xét mô hình sau:

Với mỗi cá thể,

- Nếu **asymptomatic**
    - Không **tracing** được và nó sinh ra **secondary case**
    - Các ca thể hệ thứ 2 trở đi bắt nguồn từ nó chỉ được **trace** nếu là **symptomatic**
  - Nếu **symptomatic** thì **tracing** được với xác suất  $\rho$  sau khi có **symptom**.
    - Nếu **tracing** được thì nó sẽ không sinh ra ca bệnh nào.
    - Nếu không **tracing** được thì
      - **Isolation** sau khi có **symptom** với quãng thời gian **onset-to-isolation** ngẫu nhiên.
      - Nó sinh ra thế hệ thứ 2. Số ca trong thế hệ 2 là ngẫu nhiên với trung bình cho trước. Mỗi ca trong thế hệ 2 có **time of infection** riêng, và chỉ được sinh ra nếu **time of infection** của nó nhỏ hơn **onset-to-isolation** của ca sinh ra nó.
- Họ không xét **quarantine**. **Quarantine** là **isolate** cá thể **asymptomatic**. Việt Nam áp dụng **quarantine**.

# R package ringbp [5]

# Cũng chỉ vài dòng code với package ringbp:

```
install.packages("remotes")
remotes::install_github("epiforecasts/ringbp", dependencies = TRUE)
library(ringbp)
library(ggplot2)

res <- ringbp::scenario_sim(n.sim = 10, num.initial.cases = 1,prop.asym=0,
                           prop.ascertain = 0.2, cap_cases = 4500, cap_max_days = 350,
                           r0isolated = 0, r0community = 2.5, disp.com = 0.16,
                           disp.iso = 1, delay_shape = 1.651524,
                           delay_scale = 4.287786,k = 0, quarantine = FALSE)

# Plot of weekly cases
ggplot2::ggplot(data=res, ggplot2::aes(x=week, y=cumulative, col = as.factor(sim))) +
  ggplot2::geom_line(show.legend = FALSE, alpha=0.3) +
  ggplot2::scale_y_continuous(name="Number of cases") +
  ggplot2::theme_bw()
ringbp::extinct_prob(res,cap_cases = 4500)
```

# Kết luận

- Tập chung vào câu hỏi cần trả lời để tìm phương pháp phù hợp nhất. Xuất phát từ ngữ cảnh rộng để biết hạn chế của từng phương pháp và mối liên hệ giữa các phương pháp. Xác định các khái niệm từ đầu và theo sát câu hỏi để tìm đại lượng phù hợp.
- Không nên sa đà vào việc một đại lượng có đúng trong mọi hoàn cảnh không. Không nên tập chung vào một phương pháp hay đại lượng sẵn có mà quên mất câu hỏi.
- Cần biết mô hình và phương pháp ước lượng được dùng để biết tài liệu đang nói về loại hệ số lây nhiễm nào
- Đại lượng lý thuyết khác với ước lượng của nó. Ý nghĩa lý thuyết của hệ số lây nhiễm có thể bị thay đổi khi ước lượng. Cần xem xét kỹ khi dùng  $R$  và các ước lượng của nó trực tiếp.
- Cập nhật  $R_0$  và  $T$  khi có chủng vi rút mới.
- Mô hình KM có giả thiết tổng quát hơn cho sự loại bỏ các ca bệnh, và áp dụng cho cả thời gian liên tục. Nhưng nó vẫn cần tính độc lập và đồng nhất.
- Cần một mô hình cho dữ liệu đầu vào để tính sai số. Với dữ liệu của Việt Nam, cần làm tỉ mỉ hơn như xem lại “Ngày xác nhận”, giá trị của  $T$ , .v.v.

# Tài liệu tham khảo

1. Cori, A. *et al.* (2013) 'A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics', *American Journal of Epidemiology*. Oxford Academic, 178(9), pp. 1505–1512.
2. Driessche, P. van den (2017) 'Reproduction numbers of infectious disease models', *Infectious Disease Modelling*. KeAi Publishing, 2(3), p. 288.
3. Fraser, C. (2007) 'Estimating individual and household reproduction numbers in an emerging epidemic', *PLoS ONE*, 2(8).
4. He, X. *et al.* (2020) 'Temporal dynamics in viral shedding and transmissibility of COVID-19', *Nature Medicine*. Springer Science and Business Media LLC.
5. Hellewell, J. *et al.* (2020) 'Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts.', *The Lancet. Global health*. Elsevier, 8(4), pp. e488–e496.
6. Kermack, W. O. and McKendrick, A. G. (1927) 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. The Royal Society London, 115(772), pp. 700–721.
7. Nishiura, H., Linton, N. M. and Akhmetzhanov, A. R. (2020) 'Serial interval of novel coronavirus (COVID-19) infections', *International Journal of Infectious Diseases*. Elsevier, 93, pp. 284–286.
8. Robert Koch Institut (2020) *Epidemiologisches Bulletin 17-2020: Krankheitsschwere von COVID-19, Nowcasting: Erkrankungsfälle und Reproduktionszahl*. Available at: [www.rki.de/epidbull](http://www.rki.de/epidbull)
9. Srijana (2015) *SEIR MODEL*. Available at: <https://rpubs.com/srijana/110753> (Accessed: 8 May 2020).
10. Thompson, R. N. *et al.* (2019) 'Improved inference of time-varying reproduction numbers during infectious disease outbreaks', *Epidemics*. Elsevier B.V., 29.
11. Wikipedia (no date) *Basic reproduction number*. Available at: [https://en.wikipedia.org/wiki/Basic\\_reproduction\\_number](https://en.wikipedia.org/wiki/Basic_reproduction_number) (Accessed: 25 May 2020).