

แบบจำลองโรคระบาดเบื้องต้น ๔: Statistical Estimation of the Effective Reproduction Number

ภาณุวัฒน์ เลิศสิทธิชัย

บทนำ

- ในสไลด์ชุดนี้ เราจะนำเสนอวิธีประมาณค่า Reproduction Number โดยอาศัยแบบจำลองทางสถิติแบบ Bayesian ที่ไม่ซับซ้อนมากนัก
- เป็นการประมาณค่า Effective Reproduction Number โดยใช้ข้อมูลของประเทศไทย
- เพื่อแสดงการเปลี่ยนแปลงของ **Reproduction Number** ตามเวลา และอาจทำให้หลายคนสบายใจมากขึ้น!
- ในแบบจำลองทางสถิติที่จะเสนอนี้ จะไม่มีการจำลองการติดเชื้อ ที่ครอบคลุมหรือเป็นระบบแบบ SIR (หรือ SEIR) models จึงไม่สามารถอธิบายได้ว่า ทำไม Reproduction Number จึงเปลี่ยนแปลงดังที่เห็น
- แต่มีข้อได้เปรียบคือ ไม่มีสมมติฐานตั้งต้นมากมาย จึงไม่ผิดพลาดเพราะตั้งสมมติฐานผิด
- เราเรียกการจำลองที่ไม่อาศัยสมมติฐานตั้งต้นมากนัก แต่อาศัยข้อมูลเป็นตัวชี้นำมากกว่า ว่า Empirical modeling

ทำไมต้อง Reproduction Number?

- เป็นตัวชี้วัดสำคัญของโรคติดเชื้อ ช่วยบ่งบอกถึงความยาก-ง่ายของการแพร่เชื้อ และการควบคุมเชื้อ
- เป็นตัวเลขตัวเดียว (เหมือนจะ) เข้าใจง่าย ใช้ติดตามและทำนายการระบาดของโรคได้ระดับหนึ่ง
- ประเมินค่าได้จากข้อมูลการระบาด ที่เก็บเป็นประจำ
- แต่ข้อเสียคือ มีวิธีประมาณค่าหลายวิธี และค่าที่ประมาณได้ก็อาจไม่ตรงกัน ทำให้เกิดความสับสน
- ความหมายของ Reproduction Number จึงขึ้นกับนิยามและวิธีประมาณค่าด้วย
- ในที่นี้ Reproduction Number หมายถึงจำนวนคนโดยเฉลี่ย ที่ได้รับเชื้อในปัจจุบัน ต่อการป่วยของผู้ป่วย ๑ คนในอดีต ในสังคมที่อย่างน้อยบางคนยังสามารถติดเชื้อได้
- ค่า > 1 บ่งบอกว่ามีโอกาสเกิดภาวะโรคระบาดสูง หรือการระบาดจะยังดำเนินต่อ แต่ถ้า < 1 แนวโน้มการระบาดจะน้อยมาก หรือการระบาดจะช้าลงหรืออาจไถ่ลยยุติ

Effective Reproduction Number

- **Effective Reproduction Number (R_t)** ณ เวลา หรือวันที่ t ใช้ติดตามการระบาดของโรค จึงไม่ใช่ค่าคงที่ และไม่เพียงแต่จะขึ้นกับคุณสมบัติของโรคเท่านั้น แต่จะขึ้นกับภูมิคุ้มกันของผู้ติดเชื้อ รวมถึงมาตรการต่างๆ ที่ใช้รักษา ป้องกัน และบรรเทาการระบาดของโรค อีกด้วย
- คำว่า Effective จึงช่วยบ่งบอกว่า ในเวลาหนึ่งๆ เชื้อจะแพร่ได้มากน้อยเพียงใด ในชีวิตจริงที่มีการติดเชื้อมาแล้วระยะหนึ่ง โดยรวมผลของการบรรเทาการระบาด และป้องกันโรค อยู่ด้วย
- ต่างจาก **Basic Reproduction Number (R_0)** ที่เป็นค่าคงที่โดยประมาณ โดยเสมือนเป็นตัวชี้วัดความสามารถในการแพร่ระบาดในสังคมนั้น หากไม่มีภูมิคุ้มกัน หรือมาตรการป้องกันและรักษาแต่อย่างใด จึงมีประโยชน์ในการทำนายการแพร่ระบาดในระยะต้นของโรค (และจะไม่กล่าวถึงอีกในที่นี้!)

วิธีประมาณค่า

- ใช้วิธีของ Cori *et al.*, 2014*
- ประมาณค่า Effective Reproduction Number (R_t) เป็น Posterior mean ภายใต้สถิติแบบ Bayesian
- กำหนดให้ Likelihood เป็น Poisson likelihood และ Prior distribution เป็น Gamma distribution ทำให้ Posterior distribution เป็น Gamma เช่นเดียวกัน
- จะกำหนด Poisson distribution/likelihood ก็ต้องสมมติว่า จำนวนติดเชื้อต้องน้อยกว่าจำนวนประชากรมากๆ (จะขอรายละเอียดเกี่ยวกับวิธีการประมาณค่า พอให้ผู้สนใจบางท่าน พอนึกภาพออกว่ากำลังทำอะไรอยู่ แต่สำหรับท่านอื่นๆ ไม่ต้องสนใจมากนักก็ได้!)
- ภายใต้แบบจำลองนี้ จะคำนวณได้ทั้งค่าเฉลี่ย และ 95% CI (ใน Bayesian statistics CI คือ Credible Interval) โดยอาศัยการเขียนคำสั่งใน Stata v.14

แบบจำลอง (โปรดให้ความสนใจตามสมควร!)

- **Prior distribution** เป็น Gamma distribution แบบ *Informative* อันหมายความว่า เรามีความเชื่อดั้งเดิมว่า เชื้อ COVID-19 น่าจะมีค่า Reproduction number ประมาณ 2 โดยเฉลี่ย (ก่อนนำข้อมูลมาวิเคราะห์)

$$R_{t,prior} \sim Ga(a, b) = Ga(1, 2)$$

Gamma distribution ที่มี shape parameter (a) = 1; scale parameter (b) = 2; ค่าเฉลี่ย = $ab = 2$

- **Likelihood** ของอุบัติการณ์ I_t ณ วันที่ t ที่มีการแจกแจงแบบ Poisson มี Rate parameter ($R_t \Lambda_t$) ที่กำหนดโดยข้อมูลในประเทศไทยร่วมกับข้อมูล Serial Interval distribution ของ COVID-19 ที่ได้จากประเทศจีน*

$$I_t | R_t, I_{t-1}, \dots, I_0 \sim Pois(R_t \Lambda_t) \text{ โดย } \Lambda_t = \sum_{k=1}^t I_{t-k} W_k$$

แบบจำลอง (ต่อ)

- **Posterior distribution** ของ R_t จึงเป็น Gamma distribution

$$R_{t,posterior} \sim Ga \left(a + I_t, \frac{1}{\frac{1}{b} + \Lambda_t} \right); \text{ค่าเฉลี่ย } R_{t,mean} = \frac{a + I_t}{\frac{1}{b} + \Lambda_t}$$

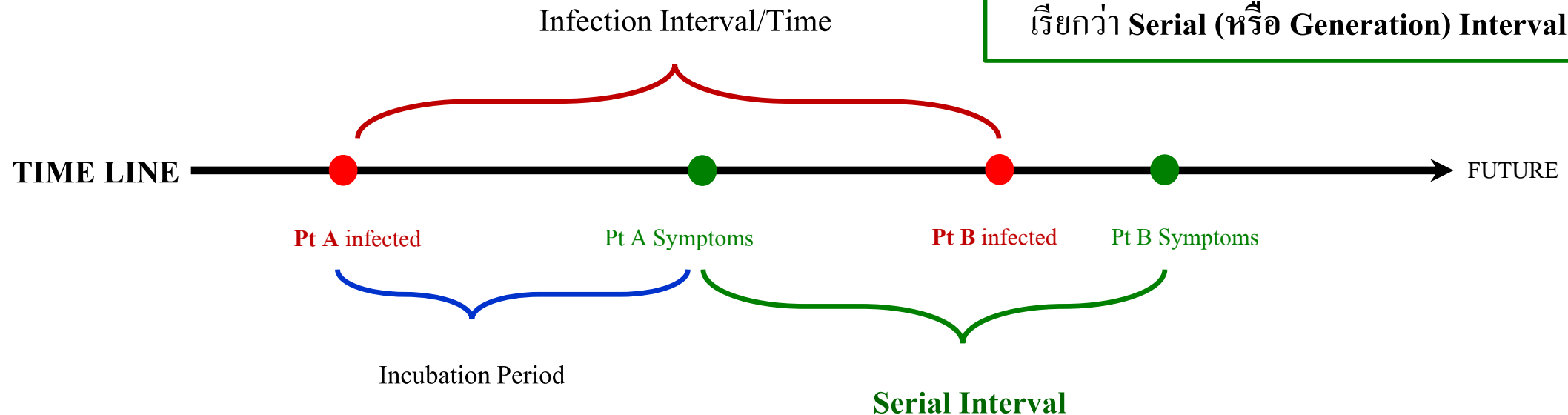
ซึ่งเราจะนำมาสร้างเป็นกราฟให้ดูต่อไป

- สังเกตว่าข้อมูลที่ใส่ในแบบจำลองประกอบด้วย I_t ซึ่งได้แก่จำนวนติดเชื้อใหม่ ณ วันที่ t และ
- ค่า W_k คือโอกาสที่คนที่ติดเชื้อและแสดงอาการ ณ วันที่หนึ่งๆ จะแพร่เชื้อให้คนอื่นอีกคนที่แสดงอาการใน k วันต่อมา อันได้มาจากข้อมูลการแจกแจง Serial Interval ของงานวิจัยจากประเทศจีน
- แต่ Serial Interval (บางคนเรียก Generation Interval) คืออะไร?

Serial Interval (อันนี้น่าสนใจและควรรู้)

ใน Time line นี้ ผู้ป่วย A แพร่เชื้อให้ B

ในที่นี้ ระยะเวลาระหว่างอาการแสดงติดเชื้อของผู้ป่วย 2 คนที่คนหนึ่งแพร่ไปยังอีกคนหนึ่งเรียกว่า **Serial (หรือ Generation) Interval**



ระยะเวลาระหว่างการติดเชื้อเรียกว่า Infection Interval ซึ่งสังเกตไม่ได้โดยตรง แต่นักวิจัยหลายท่านนิยามว่า Generation

Interval ก็คือ Infection Interval

แบบจำลองโรคระบาดเบื้องต้น ๔ : อ.นพ.ภาณุวัฒน์ เลิศสิทธิรัชย์

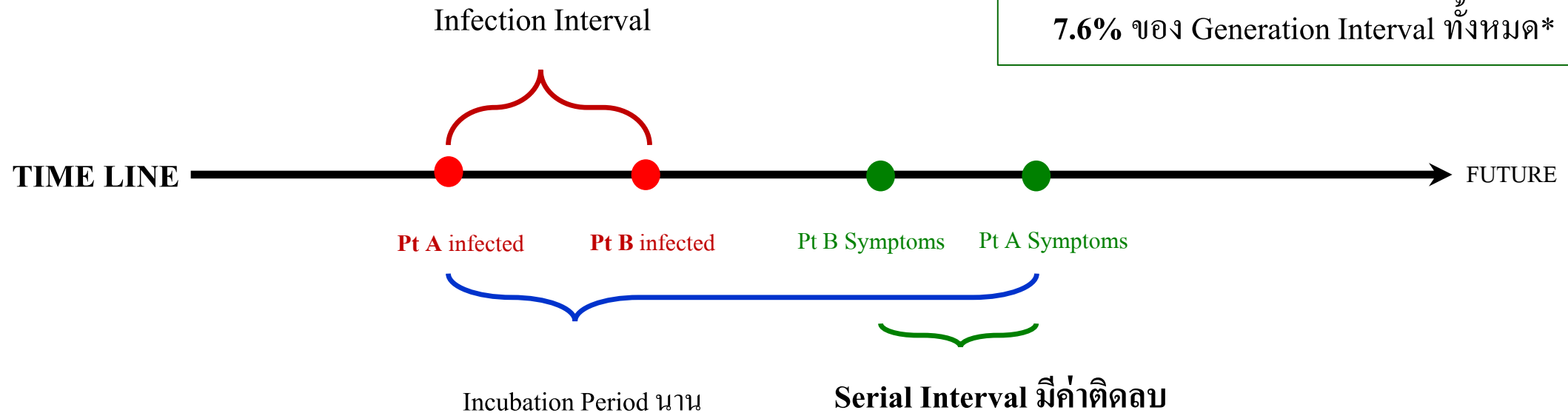
ในโรค SARS ค่า Incubation Period จะน้อยกว่า Serial Interval ดังในรูปนี้

เราสามารถเก็บข้อมูลการแจกแจง Serial Interval จากข้อมูลผู้ป่วย (ที่แพร่เชื้อถึงกัน) ได้โดยตรง

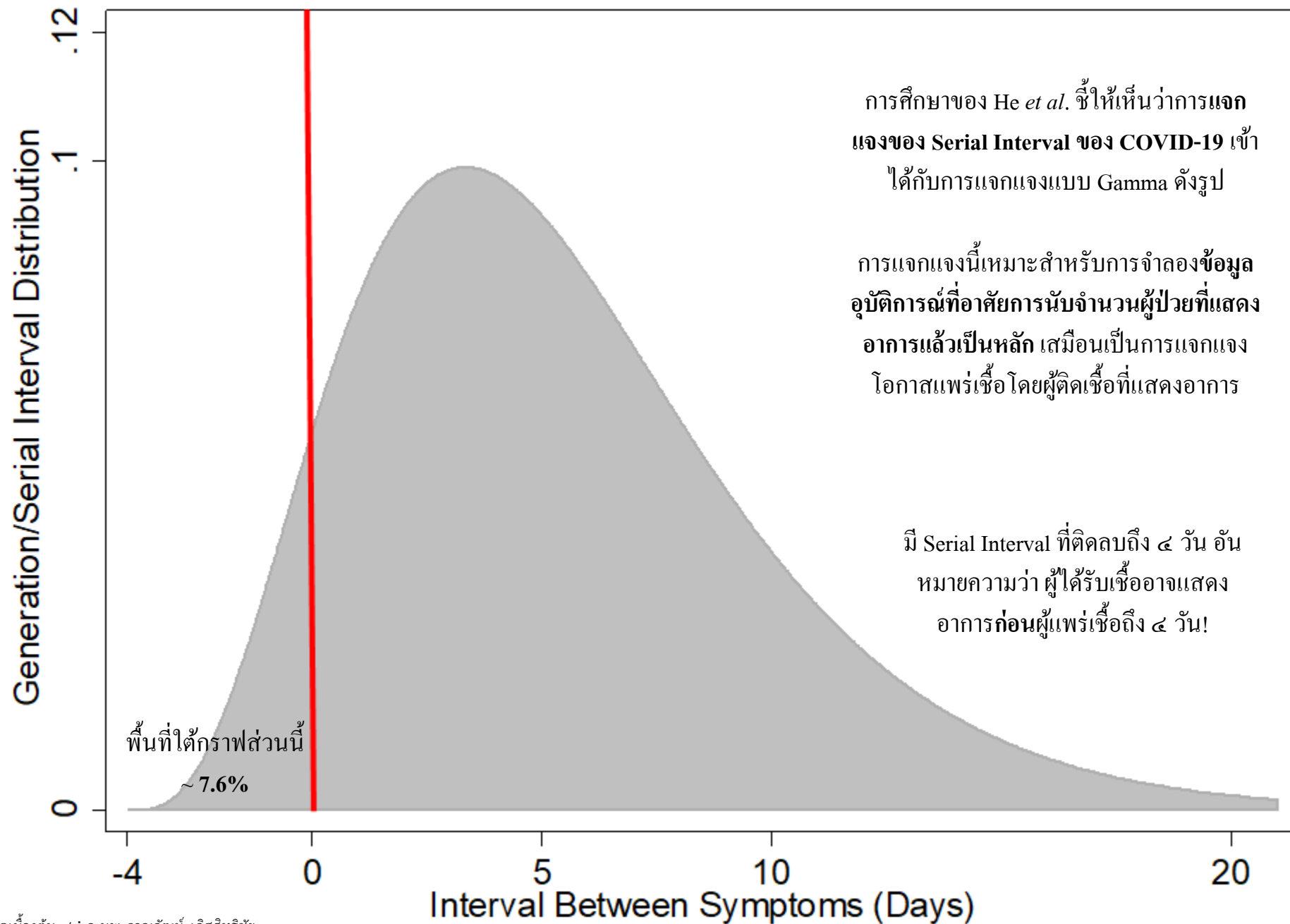
Serial Interval (ต่อ)

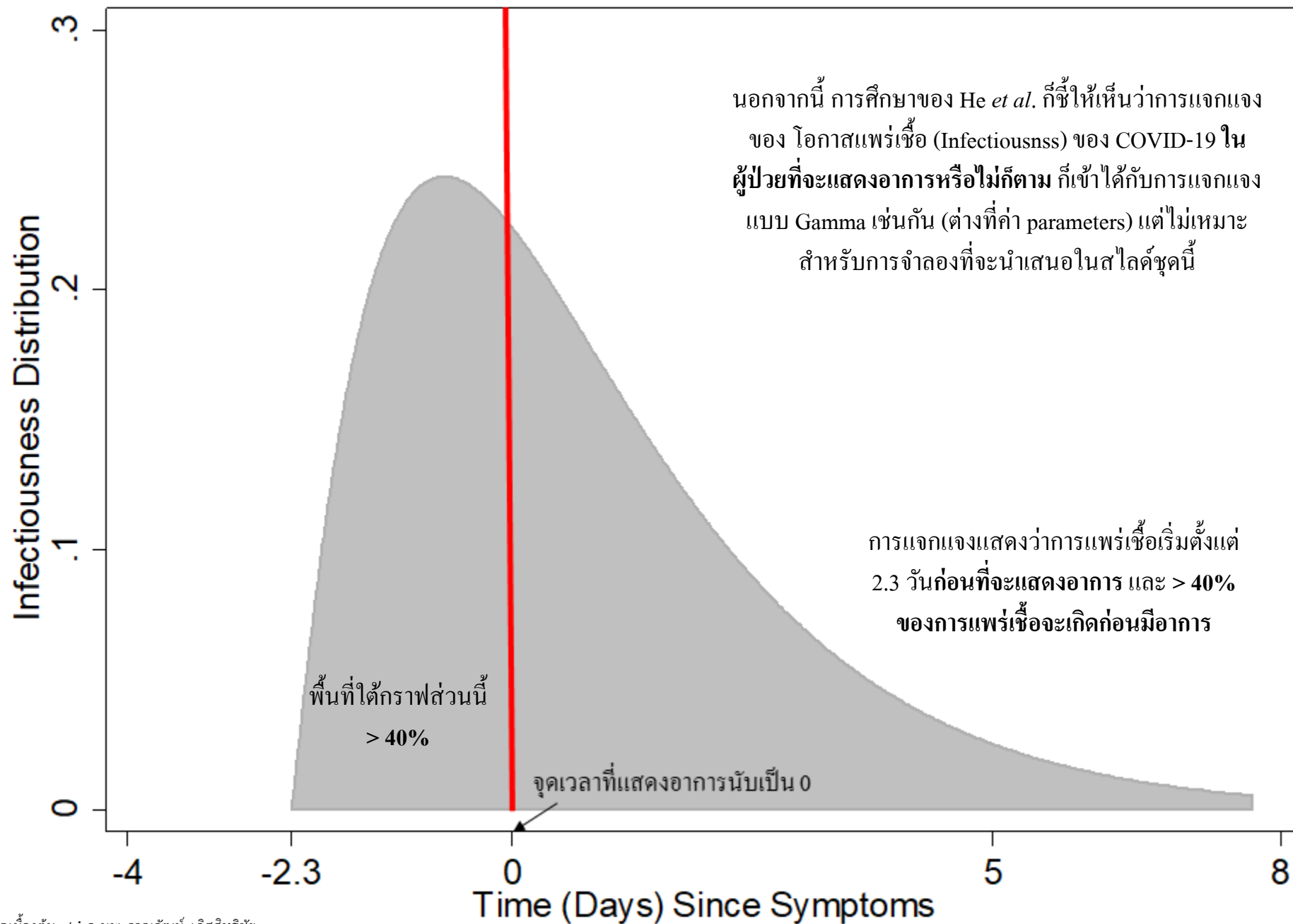
ใน Time line นี้ ผู้ป่วย A แพร่เชื้อให้ B

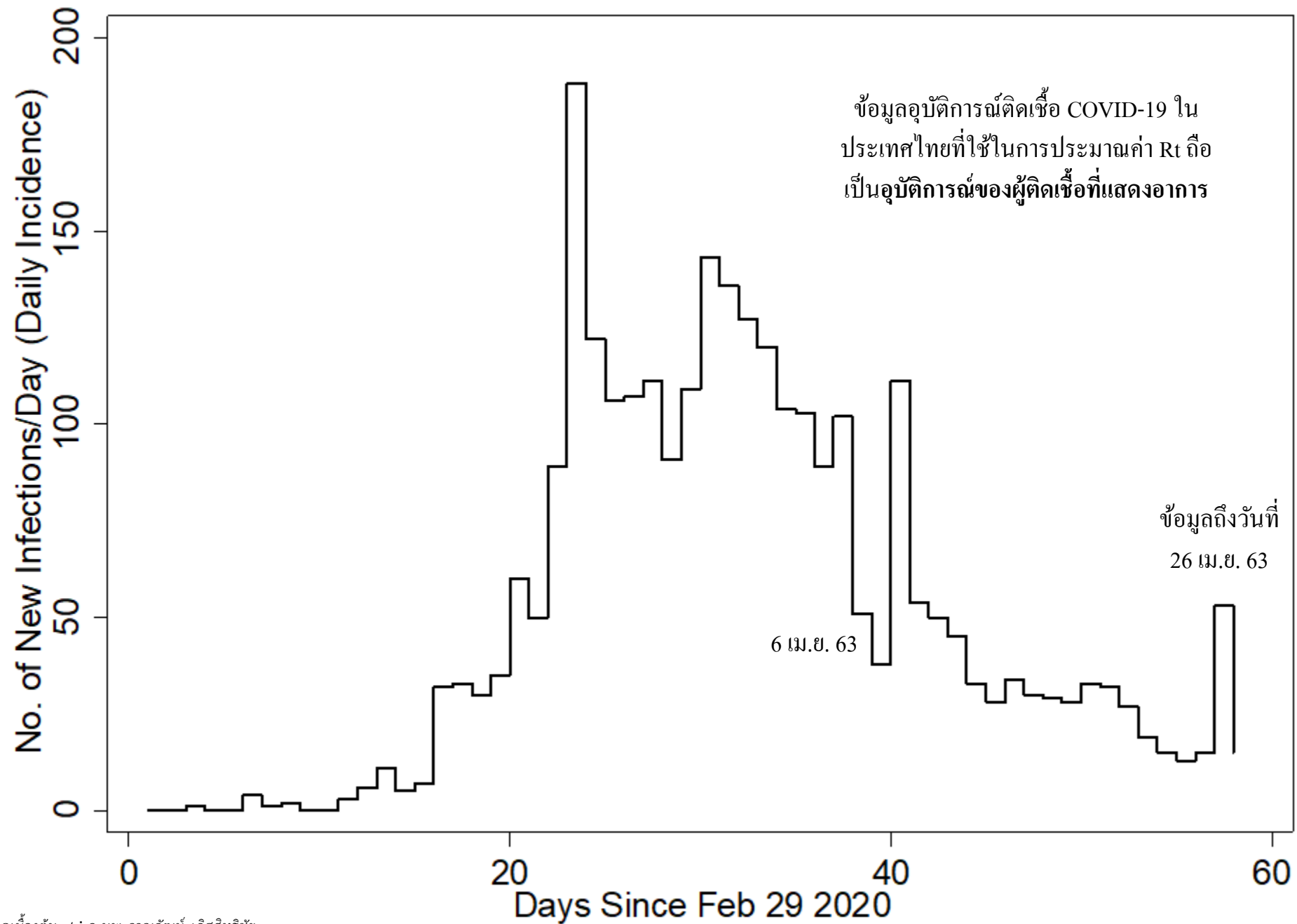
ในกรณีของ COVID-19 พบว่า เกิด *Negative* Serial Interval (อาการของ B เกิดก่อน A) ถึง 7.6% ของ Generation Interval ทั้งหมด*



*He X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med, 2020

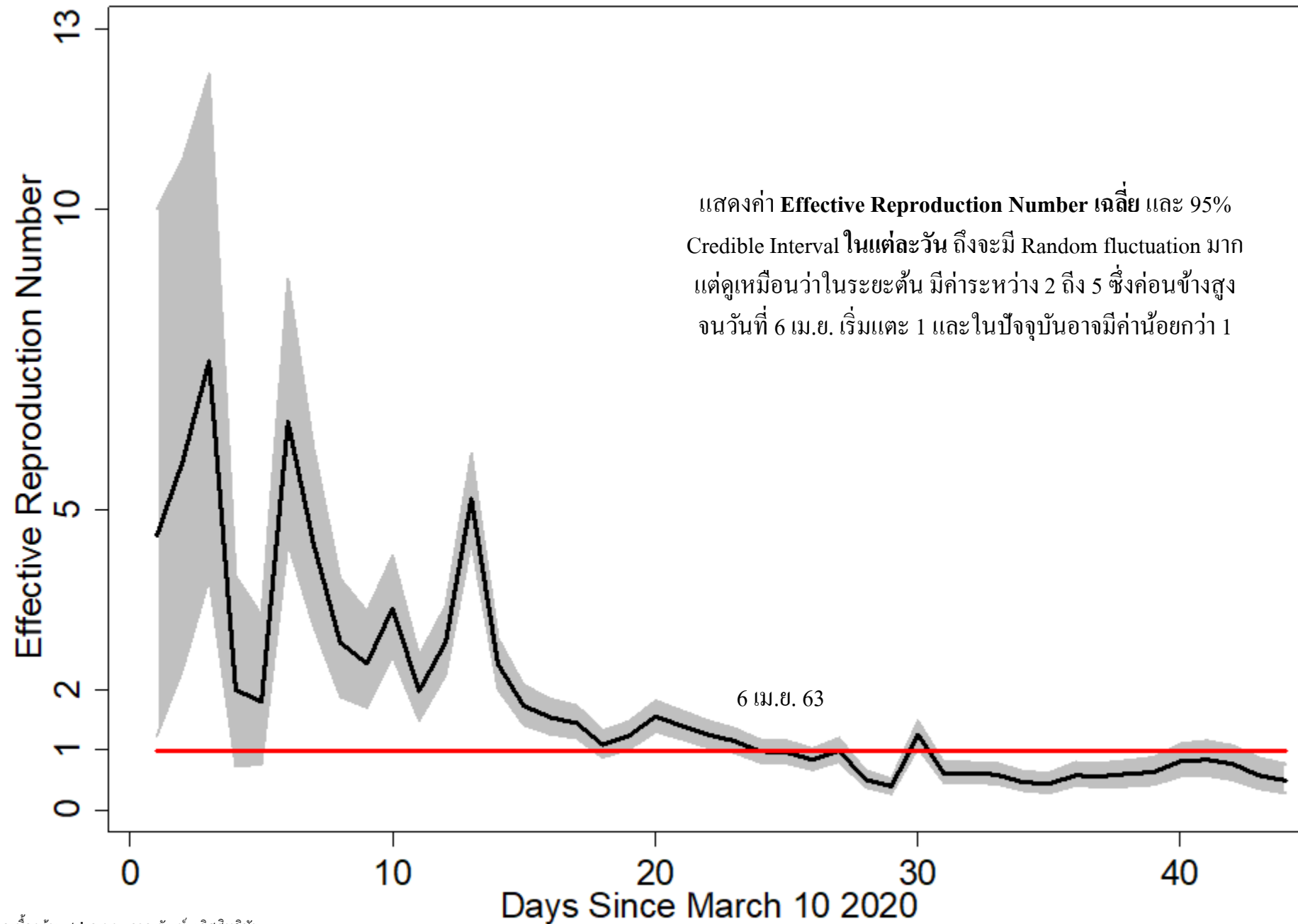


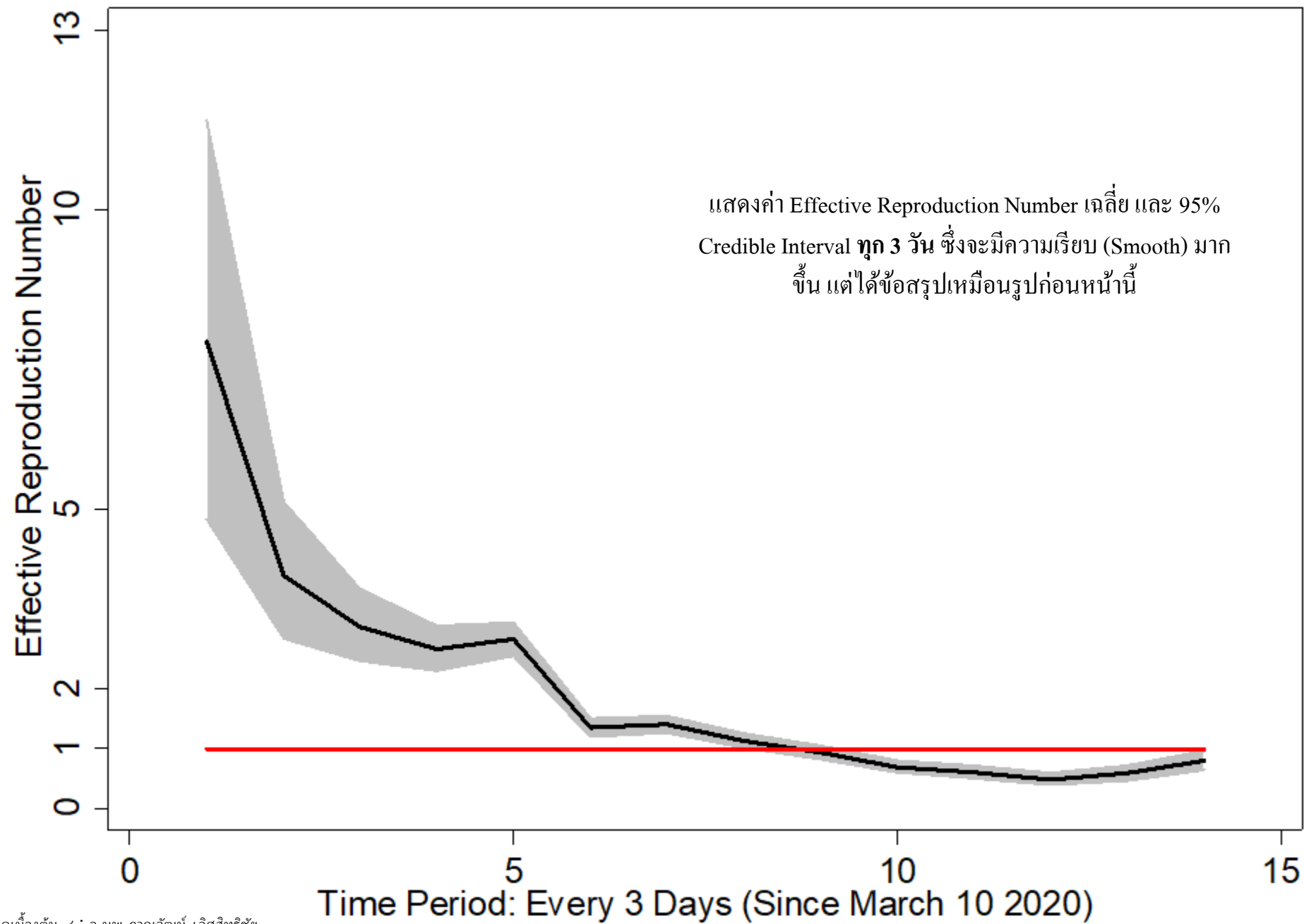




ค่าประมาณ Effective Reproduction Number เฉลี่ย

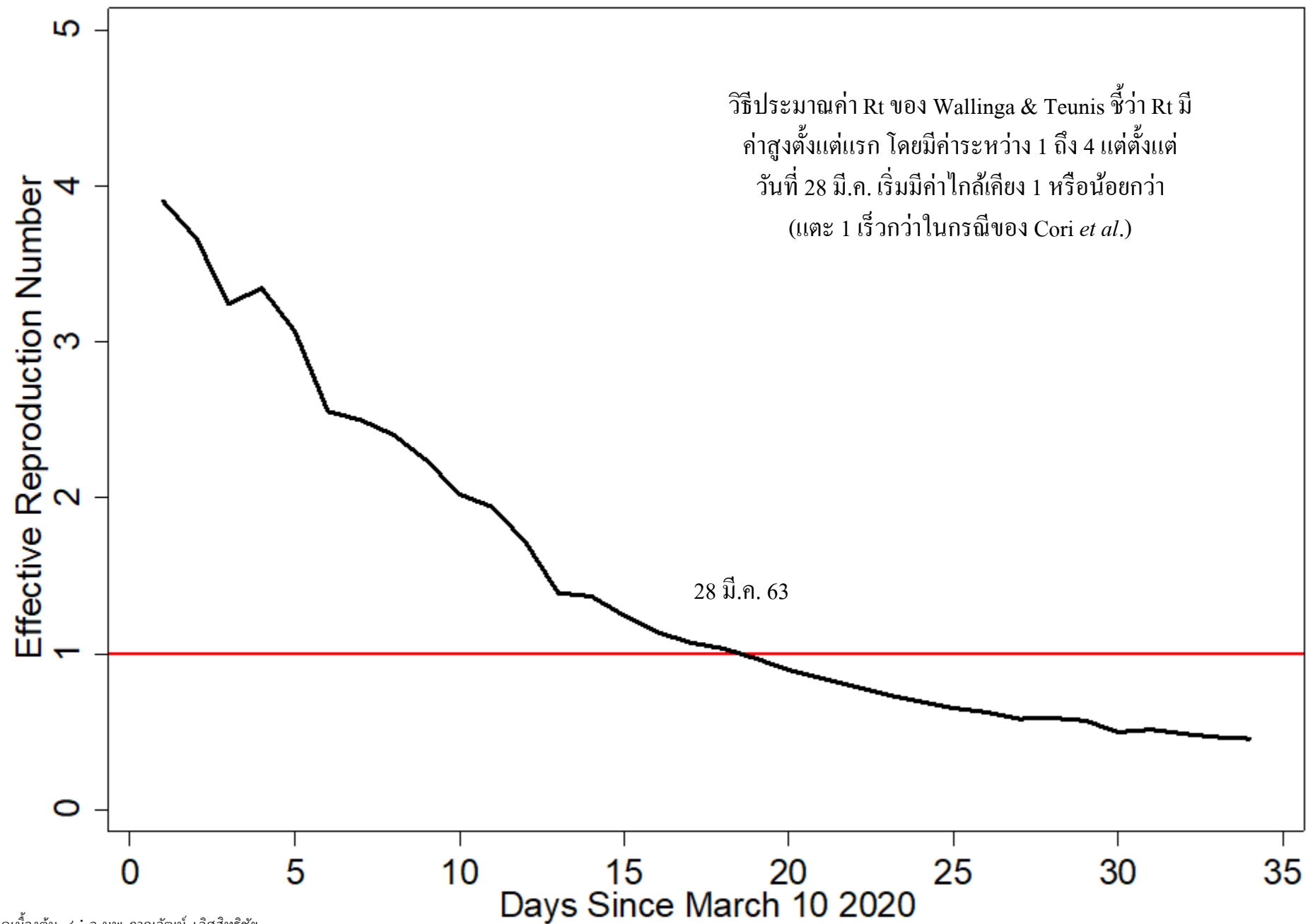
- นำข้อมูลอุบัติการณ์ I_t ข้อมูลการแจกแจง Serial Interval W_t และแบบจำลองทางสถิติที่กล่าวถึงแล้ว มา รวมกัน สามารถประมาณค่า R_t เฉลี่ย และ 95% Credible Interval ดังในสไลด์ถัดไป
- ค่า 95% CI แสดงเป็นแถบสีเทารอบค่าเฉลี่ย
- ค่า 95% CI น่าจะแคบกว่าความเป็นจริง ทำให้ดูเหมือนว่าค่าที่ประมาณได้ แม่นยำมาก
- การสร้างแบบจำลองที่มีความแปรปรวนหรือมีความไม่แน่นอนมากกว่านี้ ทำได้ไม่ยาก* แต่จะไม่กล่าวถึงในที่นี้





ค่าประมาณ Effective Reproduction Number เกลี้ย (ต่อ)

- ดูเหมือนว่า ค่า R_t จะลดลงจนเริ่มน้อยกว่า 1 และหากไม่มีการเปลี่ยนแปลงอื่น การระบาดน่าจะยุติในไม่ช้า
- แต่เราควรเปรียบเทียบการประมาณค่า R_t โดยใช้วิธีอื่น เช่นวิธีของ Wallinga และ Teunis* อันเป็นที่นิยมระดับหนึ่ง เพื่อดูว่าข้อสรุป จะเปลี่ยนแปลงหรือไม่ ดังในสไลด์ถัดไป (แต่ไม่มีค่า 95% Confidence Interval)
- การประมาณ 95% Confidence Interval ค่อนข้างยุ่งยาก จึงไม่ได้นำเสนอในรายละเอียด
- สังเกตว่าวิธีของ Wallinga & Teunis ได้ค่า R_t ไม่เหมือน (แต่คล้าย) ของ Cori *et al.* เพราะวิธีประมาณค่าต่างกัน และมีความหมายต่างกันอยู่บ้าง



ข้อจำกัดของผลวิเคราะห์ และสรุป

- ถ้าเราเชื่อถือการประมาณค่า R_t ที่นำเสนอ ก็อาจคิดว่าขณะนี้ ค่า R_t ต่ำลงมาก จนการระบาดน่าจะยุติ
- แต่อุบัติการณ์ของการติดเชื้อ COVID-19 ที่รายงานในปัจจุบัน น่าจะต่ำกว่าความเป็นจริง
- อย่างไรก็ตาม การศึกษาของทั้ง Cori *et al.* และ Wallinga & Teunis ชี้ว่า การ **Under-report** อุตบัติการณ์ มีผลกระทบต่อค่าประมาณค่า R_t ไม่มากนัก
- ดังนั้นถึงจะประมาณค่า R_t จากข้อมูลที่ปรากฏในสื่อ ก็น่าจะเชื่อถือได้ไม่มากนักน้อย แต่การ Under-report หมายความว่ายังมีผู้ป่วยจำนวนหนึ่ง ที่ยังแพร่เชื้ออยู่แต่ไม่มีอาการ
- และจะยังมีผู้ที่กลับจากต่างประเทศ ในอนาคต ที่อาจติดเชื้อแล้วและกลับมาแพร่เชื้อได้อีก
- ในขณะนี้ **Social distancing, Case identification, Isolate ผู้ติดเชื้อ** จึงยังต้องดำเนินต่อไป แต่อาจผ่อนคลายความเข้มข้นของการงดกิจกรรมทางสังคมต่างๆ ที่มีความเสี่ยงต่ำ อย่างค่อยเป็นค่อยไปได้

Technical Appendix (1)

The basic idea of Cori *et al.** was to relate R_t to the expected incidence at time, or date, t :

- $E(I_t) = R_t \Lambda_t$ where the following formula for Λ_t was used in a Stata v.14 code (available on request) adapted for partly negative-valued generation interval distributions:
- $\Lambda_t = \sum_{\substack{k=1 \\ k \neq \varepsilon+1}}^{t+\varepsilon} I_{t+\varepsilon+1-k} W_k$; where ε is the maximum negative generation interval (here, $\varepsilon = 4$ days) and I_t starts at I_1 ;
- W_k was calculated using a formula in Cori *et al.*, which can also be replaced by a simpler discretized version:
- $W_k = F_{Ga(a,b)}(k + 0.5) - F_{Ga(a,b)}(k - 0.5)$, where $F_{Ga(a,b)}(k)$ is the cumulative distribution of a Gamma distribution with parameters a , b (*with very similar results*);
- And the Generation Interval Density was $Ga(4.5, 2.1)$.

Technical Appendix (2)

The idea of Wallinga and Teunis* was to estimate R_t at time t_j by (N = total number of time points):

- $R_j = \sum_{i=1, i \neq j}^N \frac{W(t_i - t_j)}{\sum_{k=1, k \neq j}^N W(t_i - t_k)}$; where $W(t_i - t_j)$ is the Generation Interval Distribution for the interval $(t_i - t_j)$: note that the sum, **in contrast to Cori et al.**, is over all past & future observations;
- This was implemented in a Stata v.14 code (available on request) adapted to the daily incidence data and the discretized distribution (*of partly negative-valued intervals*) as per Cori et al., (as well as the simpler discretized version stated previously) by**:
- $R_j = \sum_{i=1, i \neq j}^{N-\varepsilon} \frac{I_i W_{i+\varepsilon+1-j}}{\sum_{k=1, k \neq \varepsilon+1}^{i+\varepsilon} I_{i+\varepsilon+1-k} W_k}$; however, there was instability at both extreme ends of estimates.

*Wallinga J & Teunis P. Different epidemic curves for SARS reveal similar impacts of control measures. Am J Epidemiol, 2004