

実効再生産数とその周辺

2020年5月12日

西浦 博

北海道大学

nishiurah@med.hokudai.ac.jp

<https://github.com/contactmodel/COVID19-Japan-Reff>

研究員

Sungmok Jung

Andrei Akhmetzhanov

水本憲治

小林鉄郎

吉井啓太

北大チーム、東北大チームの皆さん

Disclaimer

本日の内容は、高校理系数学以上の数理的バックグラウンドを想定して話させていただきます。数式の解釈に関しても努力して一緒に学ぶことを想定します。

本発表は発表者が専門家個人として発表するものであり、現在参画する厚生労働省クラスター対策班の意を反映したものではありません。

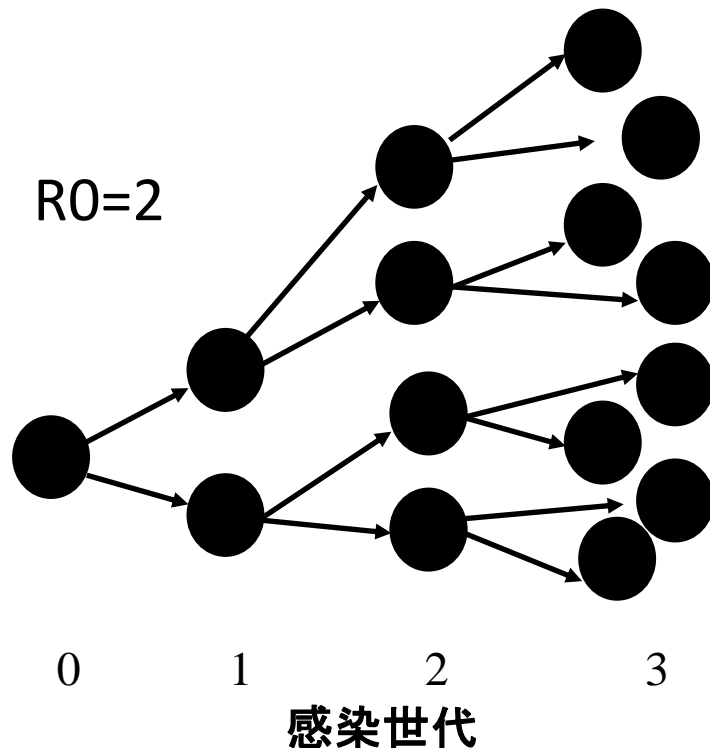
専門家として個人的にデータ収集をし、また、一部担当部署に許可を得た全国データを共有します。

<https://github.com/contactmodel/COVID19-Japan-Reff>

基本再生産数 (Basic reproduction number; R_0)

$$i(t) = \int_0^{\infty} A(\tau) i(t - \tau) d\tau$$

$$R_0 = \int_0^{\infty} A(\tau) d\tau$$



R_0 (基本再生産数) : **全ての者が感受性を有する集団**において1人の感染者が生み出す2次感染者数の平均値

流行の増殖期に入ってから

観察データをモデル化する立場から：

R_0 は本質的に**数理的**。

確率的ゆらぎ = > 精密な推定が困難

実践性として定量化しやすいものが必要。

- － 全ての者が感受性を有する
- － 感染者数増殖に線形の近似をする
- － その持つ意味は、背景にある個体群動態のシステムに依存する

Rn

Net reproductive ratio (人口学)

Net salary (実収入額)

Net time (マラソン)

「実際のところ、どうなのか」 = > 実効再生
産数 (effective reproduction number)

Non $R_0 \Rightarrow R$ (effective...)

Several non-mathematical definitions.

A: あるカレンダー時刻 t で起こっている2次感染数の1人あたり平均値.

B: あるカレンダー時刻 t で感染した1人がその後の経過で生み出す2次感染者数の平均値

C: 罹患率—有病割合比などから推定される1人あたりの2次感染者数 (actual reproduction numberとか)

D: 予防接種など流行対策下での1人当たりが生み出す2次感染者数

再生産方程式

$$\underbrace{i(t)}_{\text{時刻}t\text{における新規感}} = \int_0^\infty \underbrace{A(t, \tau)}_{\text{時刻}t\text{で感染源の感}} \underbrace{i(t - \tau)}_{\text{過去（時刻}t-\tau\text{）の新}} d\tau$$

時刻 t における新規感
染者数

時刻 t で感染源の感
染後時刻 τ にお
いて1人当たり
の2次感染率

過去（時刻 $t-\tau$ ）の新
規感染者

t : カレンダー時刻

τ : 感染齢（感染後に経過した時刻）

$$\frac{dX}{dt} = -\beta XY$$

$$\frac{dY}{dt} = \beta XY - \gamma Y \quad \rightarrow$$

$$\frac{dZ}{dt} = \gamma Y$$

$$\frac{dX}{dt} = -X \int_0^t \beta(s) Y(t-s, s) ds$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) Y(t, s) = -\gamma(s) Y(t, s)$$

$$\frac{dZ}{dt} = \int_0^t \gamma(s) Y(t, s) ds$$

$$Y(t, 0) = X \int_0^t \beta(s) Y(t-s, s) ds$$

$$i(t) = X \int_0^t \beta(s) Y(t-s, s) ds$$

A: あるカレンダー時刻 t で起きている2次感染数の1人あたり平均値.

$$R_1(t) = \int_0^{\infty} A(t, \tau) d\tau \left(= \frac{S(t)}{N} R_0 \right)$$

ピリオド合計特殊出生率に相当
=>流行対策の評価に適する

B: あるカレンダー時刻 t で感染した1人がその後の経過で生み出す2次感染者数の平均値

$$R_2(t) = \int_0^{\infty} A(t + \tau, \tau) d\tau$$
$$\left(= \frac{\beta}{N} \int_0^{\infty} S(t + \tau) \exp(-\gamma\tau) d\tau \right)$$

コホート合計特殊出生率に相当

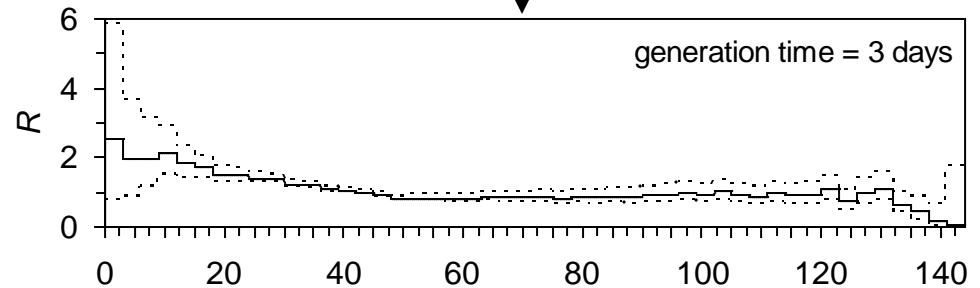
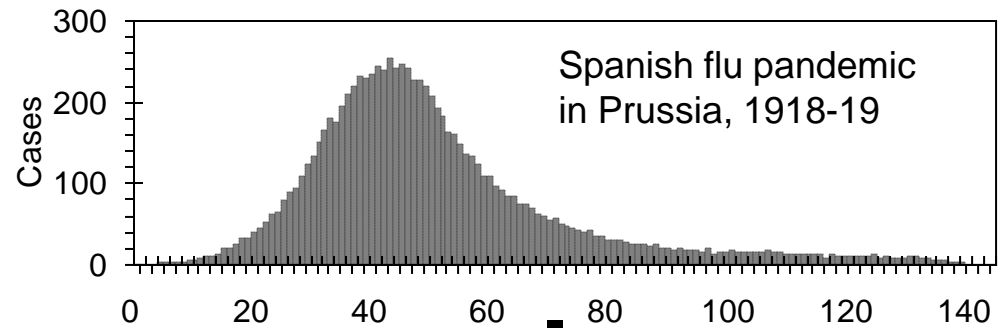
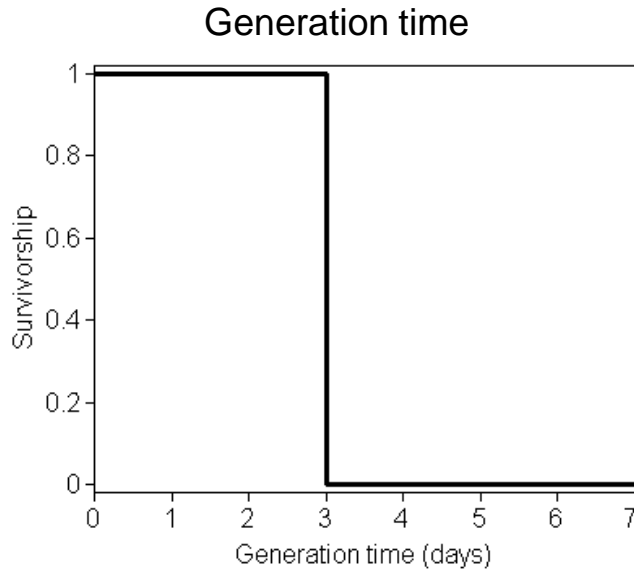
Fraser C. PLoS ONE 2007; 2: e758.

Alho & Spencer. *Statistical Demography and Forecasting*. 2005 (Springer)

$R_t < 1$

- ・「新規感染者数が減っている」
- ・どれくらいの程度で減っている（小さいほど早く流行が終息する）
- ・落とし穴はある（ $R_t < 1$ でも安定するような系）

単純化されたケース



$$E(R_n) = \frac{i_{n+1}}{i_n}$$

仮に世代別の感染者数がわかった場合
=> 当該の元世代の実効再生産数

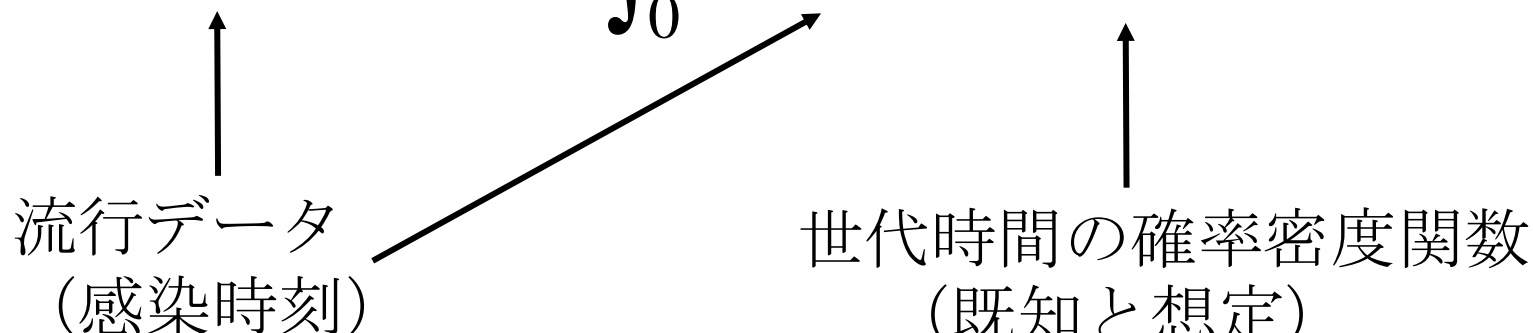
Dailyのデータにどう対応しているか

$$i(t) = \int_0^{\infty} A(t, \tau) i(t - \tau) d\tau$$

$$A(t, \tau) = R(t) g(\tau)$$

$$i(t) = R(t) \int_0^{\infty} i(t - \tau) g(\tau) d\tau$$

流行データ
(感染時刻)

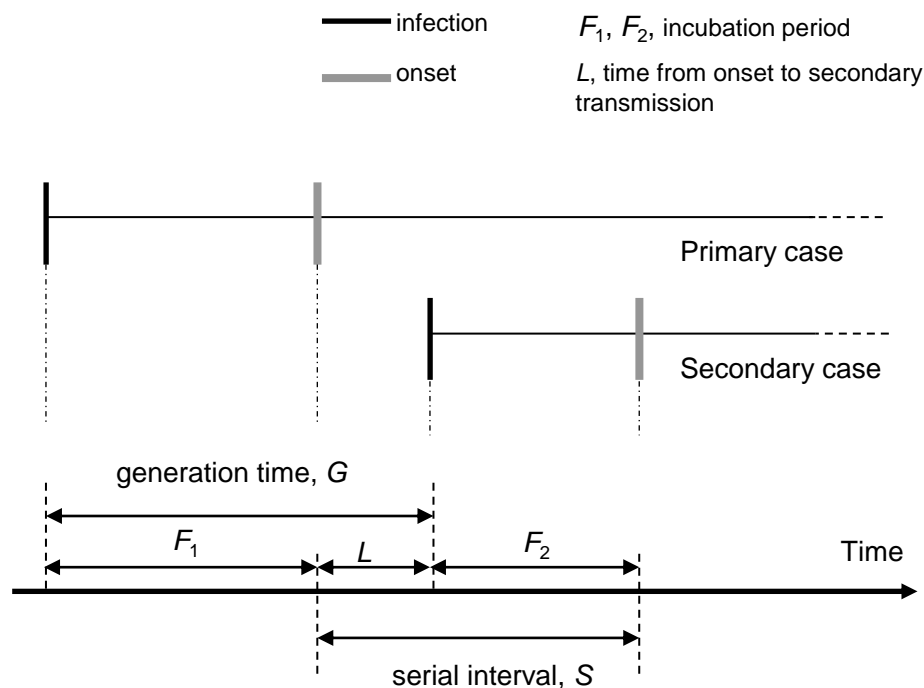


世代時間の確率密度関数
(既知と想定)

世代時間 < = 明示的にまだ得られてない 間、発病間隔と同じと想定

Serial interval (発病間隔) : 感染源の発病から2次感染者の発病までに要する時間
(平均4.8日 ; Nishiura et al. 2020)

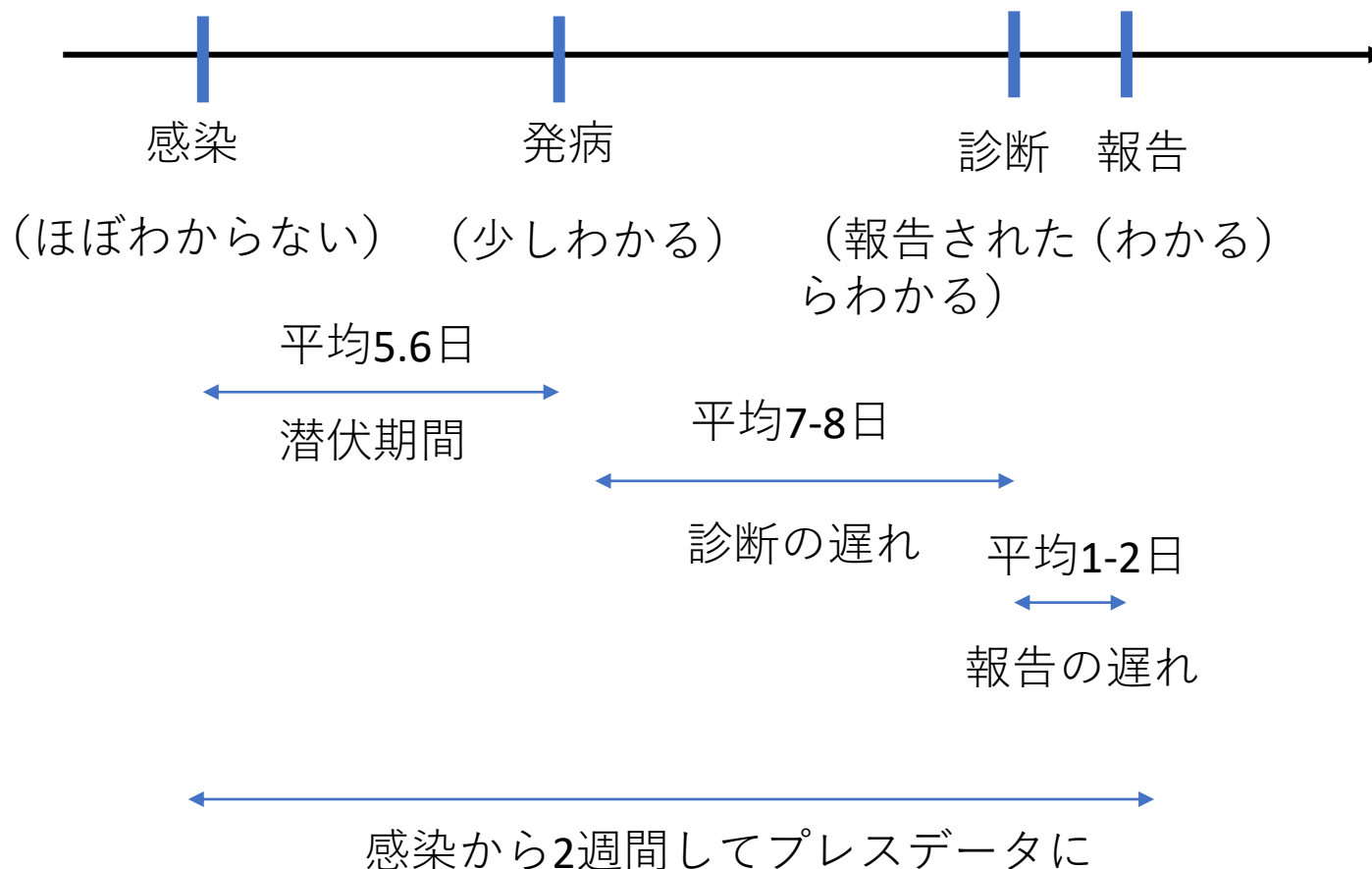
Generation time (世代時間) : 感染源の感染から2次感染者の感染までに要する時間
(ちなみに、ODEで書かれる最も単純なSIRモデルでは平均世代時間 = > 平均感染性期間)



Nishiura H, et al. Int J Infect Dis 2020; 93:284-286

Klinkenberg D, Nishiura H. J Theor Biol 2011; 284:52-60.

「患者」の観察データについて



逆計算 (Backcalculation)

東京の場合

3月25日 都知事の外出自粛呼びかけ

4月7日 政府の緊急事態宣言

5月6日 政府の緊急事態宣言延長

流行1日目: $c_1 = i_1 f_0$

流行2日目: $c_2 = i_2 f_0 + i_1 f_1$

流行3日目: $c_3 = i_3 f_0 + i_2 f_1 + i_1 f_2$

流行4日目 $c_4 = i_4 f_0 + i_3 f_1 + i_2 f_2 + i_1 f_3$

畳み込み
$$c(t) = \int_0^t f(u) i(t-u) du$$

ノンパラメトリック推定(の前まで)

$$c_t = \sum_{u=1}^t f_u i_{t-u}$$

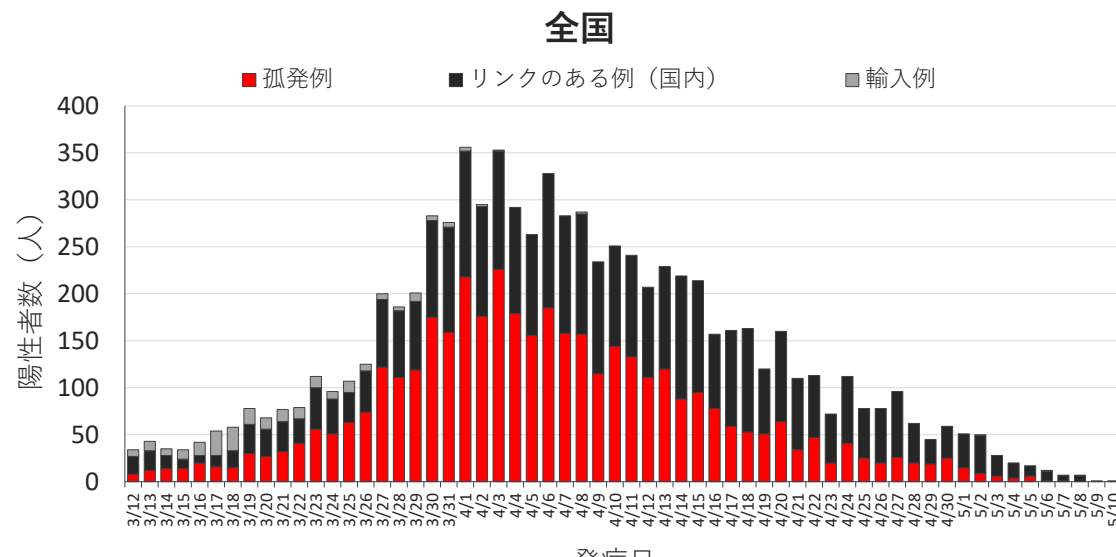
i_t が非定常ポアソン過程に従って生成されているとすると, 発病者 c_t もポアソン分布に従う. 発病者を基に感染者を推定するための尤度は以下に比例する

$$\prod_{t=1}^T \left(\sum_{u=1}^t f_u i_{t-u} \right)^{r_t} \exp \left(- \sum_{u=1}^t f_u i_{t-u} \right)$$

Rのパッケージ *surveillance* を使用

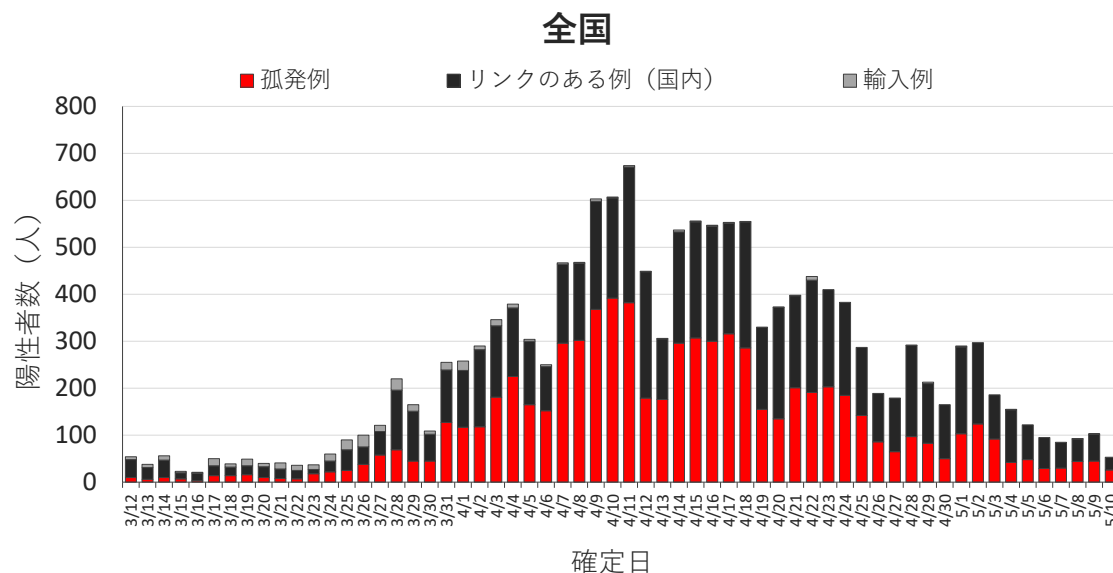
発病時刻がわからない場合：診断時刻＝＞発病時刻にした上で
全感染者で分析

日本のデータが包含する問題



- ・ 発病時刻が既知 (n=8811)、確定時刻はn=15520で既知。約200名はどちらもわからず。

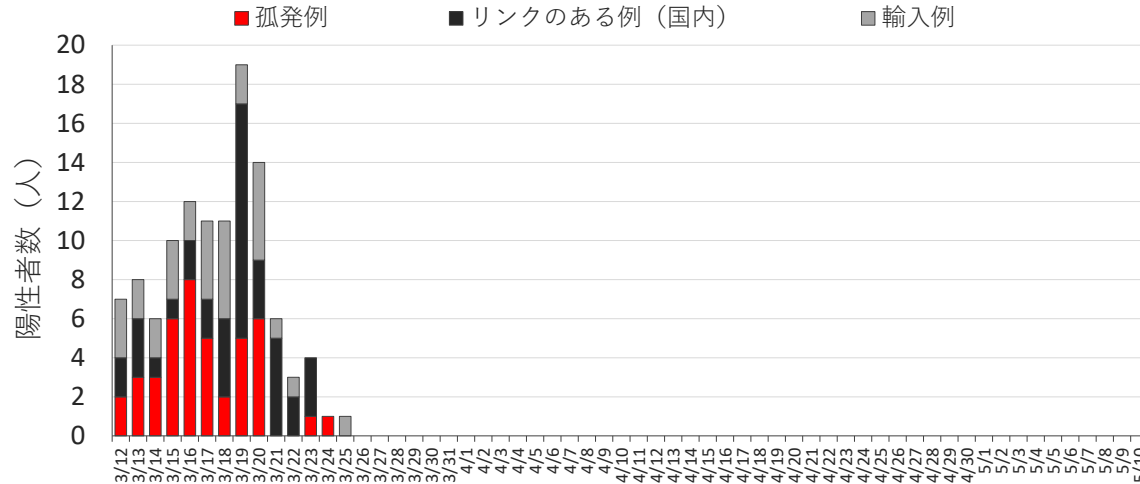
- ・ 特に東京で発病時刻未知が多数（公表ベース）



- ・ 報告の遅れ（診断 = > 報告）。数日の場合から1週間を超える場合が稀にあり。時刻とともに変動

公表に基づく専門家データ

東京

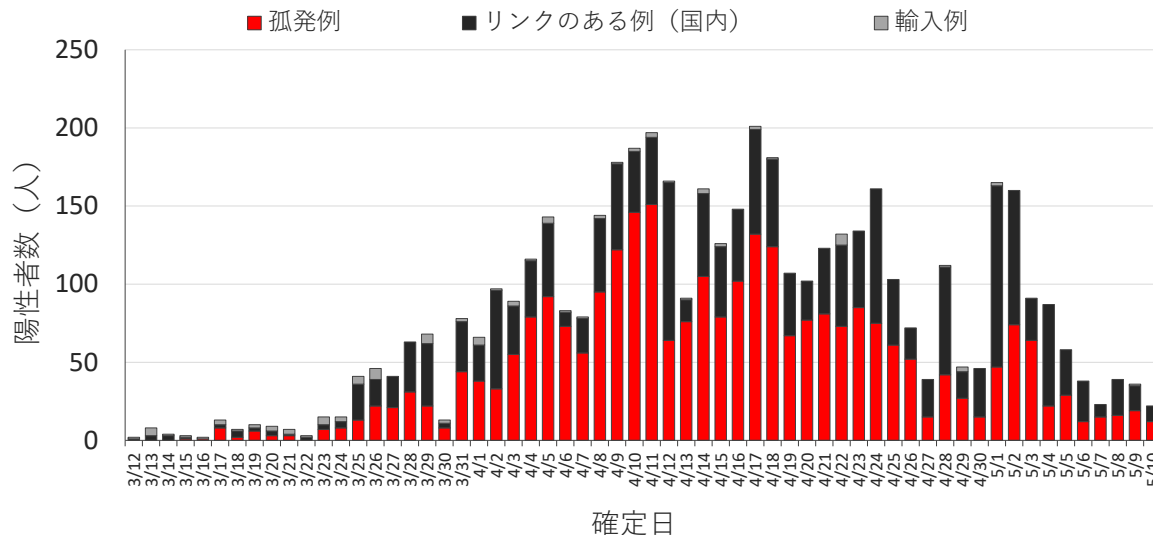


・ 専門家データ（ボランティアの力を借りて都道府県プレスリリースから構築）⇒ 東京データ抜け

・ 今回：構築データに東京都から許可を得て統合した全国データを共有（発病時刻ありOrなければ診断時刻）

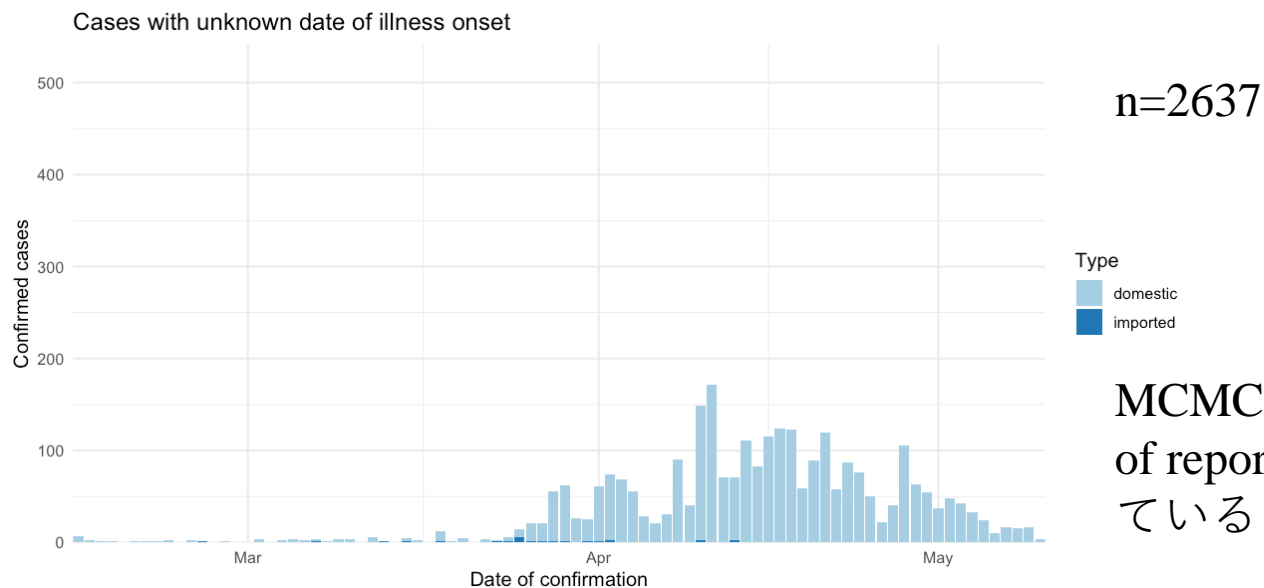
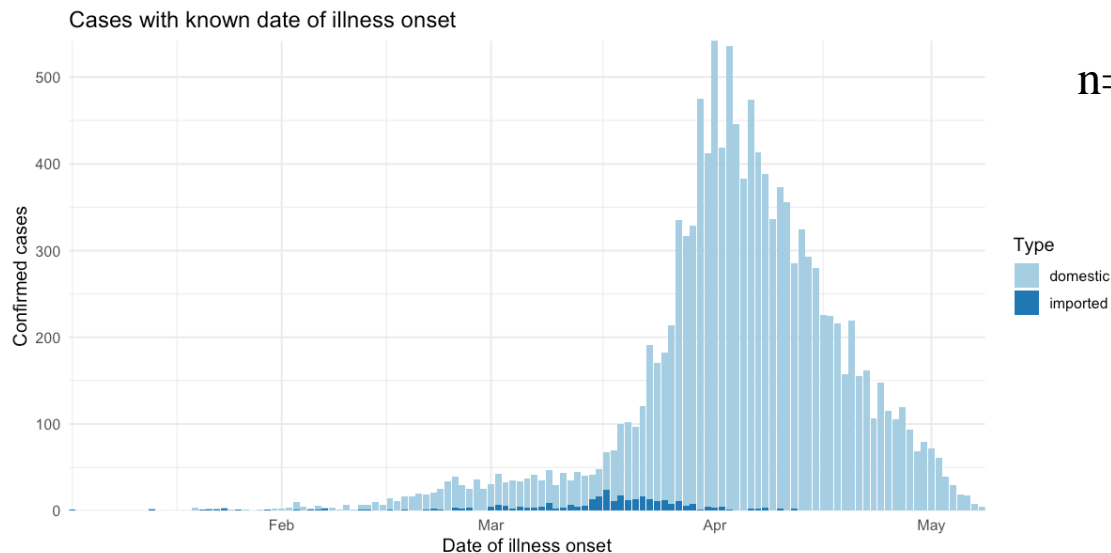
・ データは専門家でなく対策本部に管理部署あり。また、国立感染症研究所でサーベイランス運営

東京



<https://github.com/contactmodel/COVID19-Japan-Reff>

<https://nbviewer.jupyter.org/github/contactmodel/COVID19-Japan-Reff/blob/master/scripts/A.%20Showing%20the%20epidemic%20curve.ipynb>



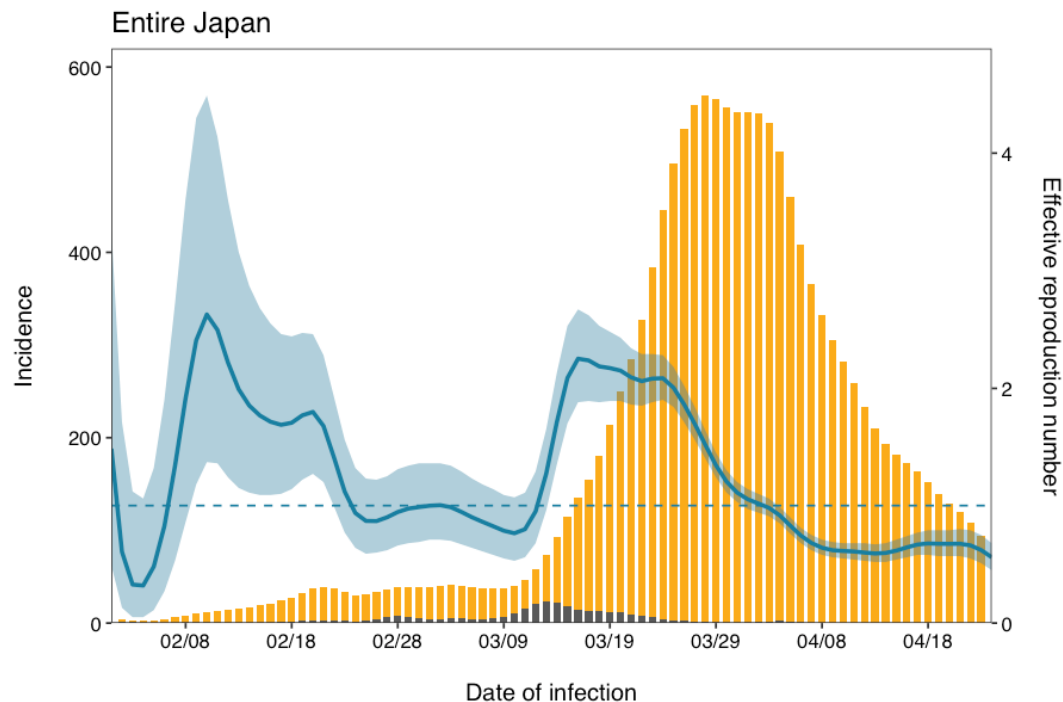
R(t)の推定へ

$$i(t) = R(t) \int_0^{\infty} i(t - \tau) g(\tau) d\tau$$

$$E(i_{\text{domestic}}(t)) = R_t \sum_{\tau=1}^{t-1} i_{\text{total}}(t - \tau) g(\tau) \frac{F(T - t)}{F(T - t + \tau)}$$

$$\begin{aligned} & L(R_t; C_{\text{domestic}}(t)) \\ &= \prod_{t=1}^T \frac{\exp(-E(i_{\text{domestic}}(t))) (E(i_{\text{domestic}}(t)))^{i_{\text{domestic}}(t)}}{i_{\text{domestic}}(t)!} \end{aligned}$$

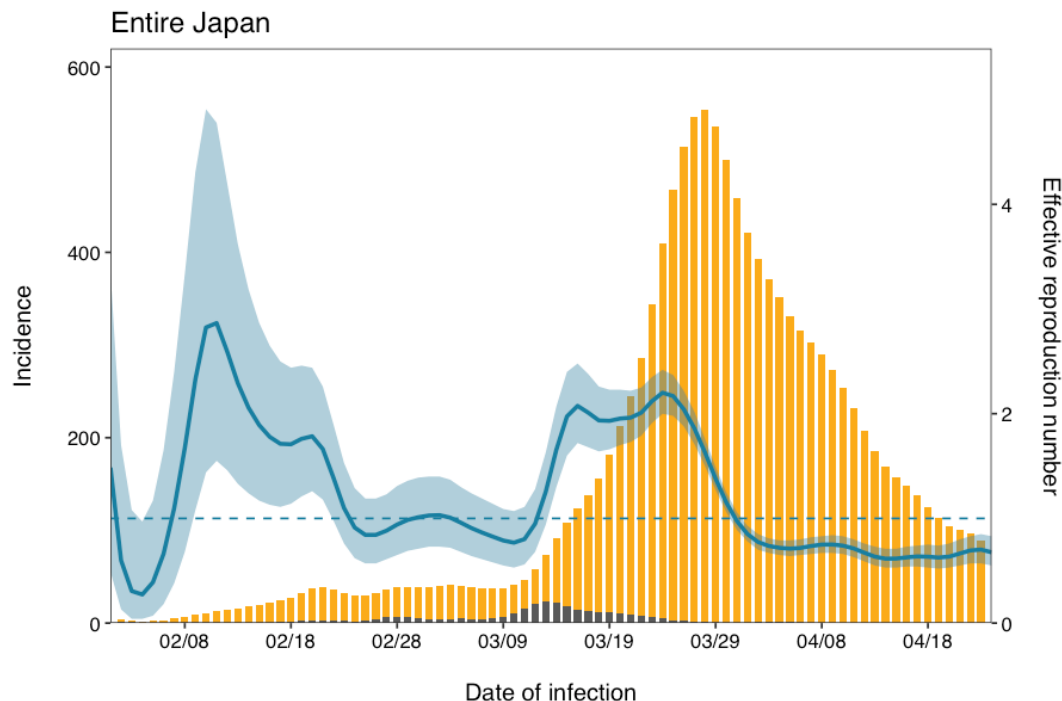
MLE=> Profile likelihood 95% CI or MCMC



上：発病日付のわからない者を除いた場合

下：発病日付がわからない者は、確定日付を基に発病日を逆計算した上で合わせて、その後、感染時刻を推定

「定性的に類似の傾向」



最近の日付（4月26日）で Truncate（感染から報告まで18日間が 80 percentile）：時間に依存する報告の遅れの影響は免れない。。

リアルタイム性に乏しい

改善を要するがこれまでに対応 できていない件

1. Imputation
2. Heterogeneity (Age)
3. Backcalculation<=Contact patterns

$$i(t) = R(t) \int_0^{\infty} i(t - \tau) g(\tau) d\tau$$

$$\frac{i(t)}{q(t)} = R(t) \int_0^{\infty} \frac{i(t - \tau)}{q(t - \tau)} g(\tau) d\tau$$

関連して、真摯に説明しないといけない件

R(t)を使っていたのが、(いつのまにか) Doubling Timeに

$$C(t) = C(0)2^{\frac{t}{T_d}}$$

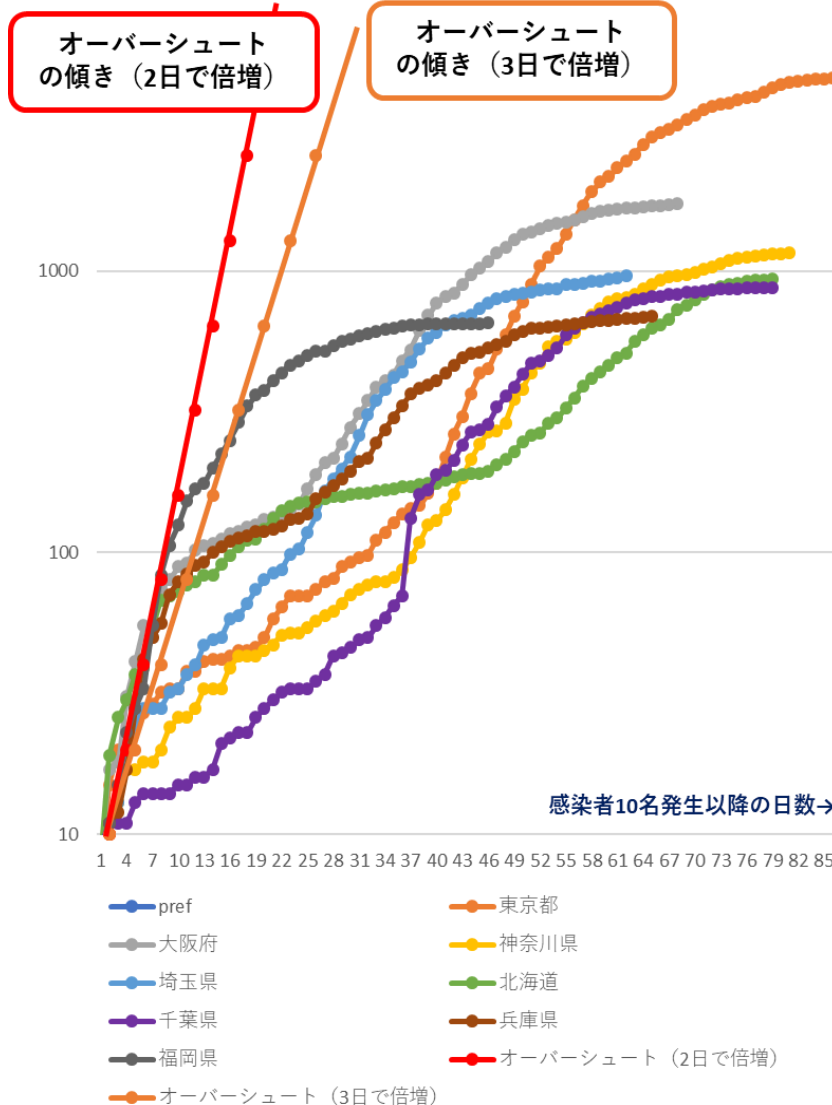
3月：欧州・米国で次々と2-3日に累積患者数が倍増

研究あり：**Backward**に政策評価をする（約2週前の**Reflection**でしかない）
Preciseな再生産数の推定値よりも、**Doubling time**をみて欧米の爆発的な感染者数の増加の2-3日に1回を見逃さないことが重要

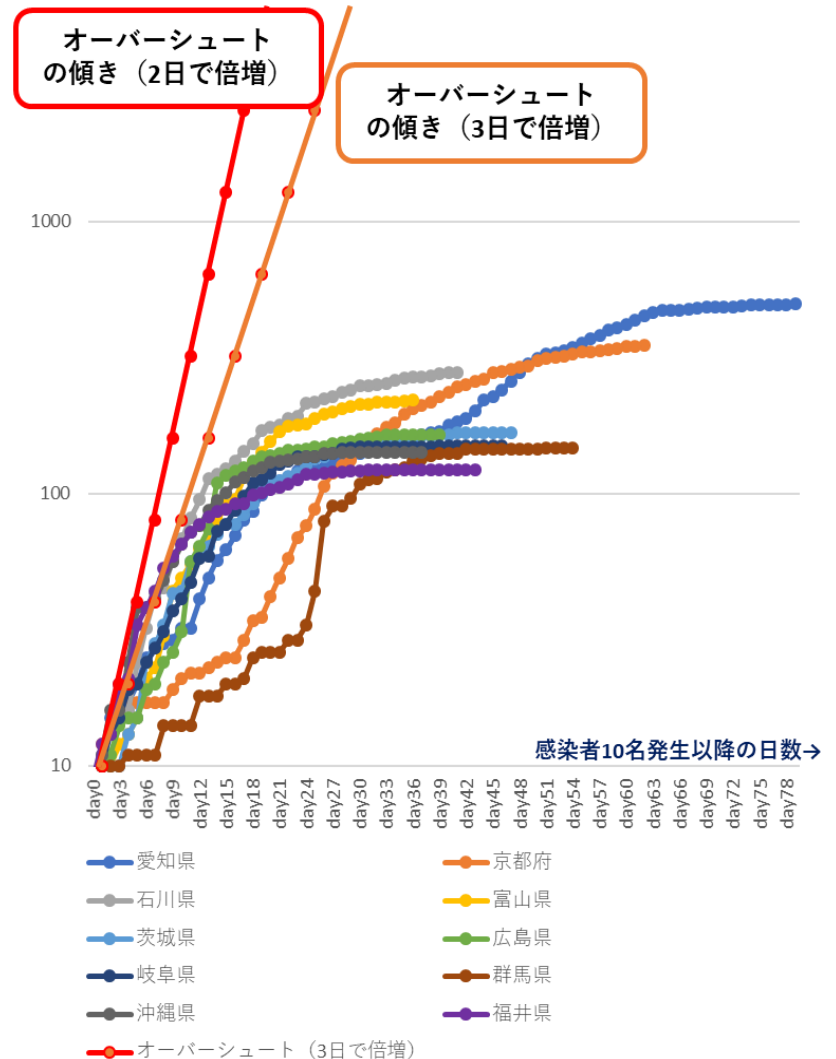
今後：患者数の増加時に2-3日に倍増でないことを確認するために有用

倍加時間の続き (2通りの計算)

累積感染者500名以上の都道府県



累積感染者100～500名の都道府県



今後 $R(t)$ については、どうして いくべきか（私見）

- ・ ダッシュボード（推定の実績のある企業の協力）
⇒ 都道府県別で R_t がアップデートされる体制
- ・ 対案（より平易な計算をする）
報告日に基づく計算によって簡易的に評価

对案: The earliest concept of $R(t)$

pp. 501

variations in the course of the epidemic. THE AUTHOR (1928) has used the measure

$$\text{Ratio of dispersibility} = \frac{\text{cases in period B}}{\text{cases in period A}} \dots\dots\dots (i)$$

ratio of cases
for adjacent period

pp. 502

Influenza in the French Army 1918-1919
Cases (192 286) and Fatality Rates for each Month, expressed as
Percentages above & below the Monthly Mean of the whole period —

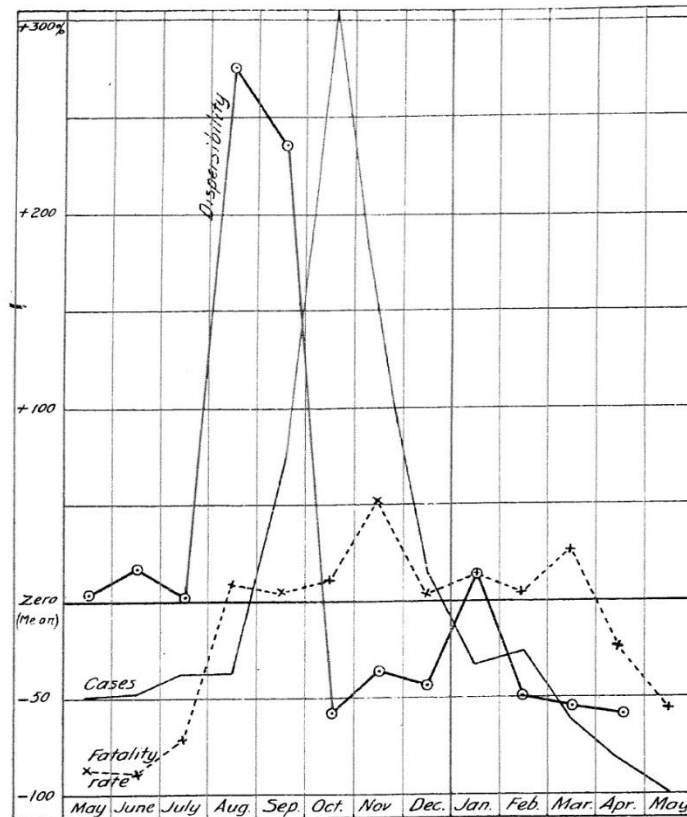


Fig. 52. Index of dispersibility, incidence and fatality of influenza in 1918.
(Constructed from figures in Delater, "La Grippe dans la Nation armée,"
Rev. d'Hygiène, xiv, 5, 406.)

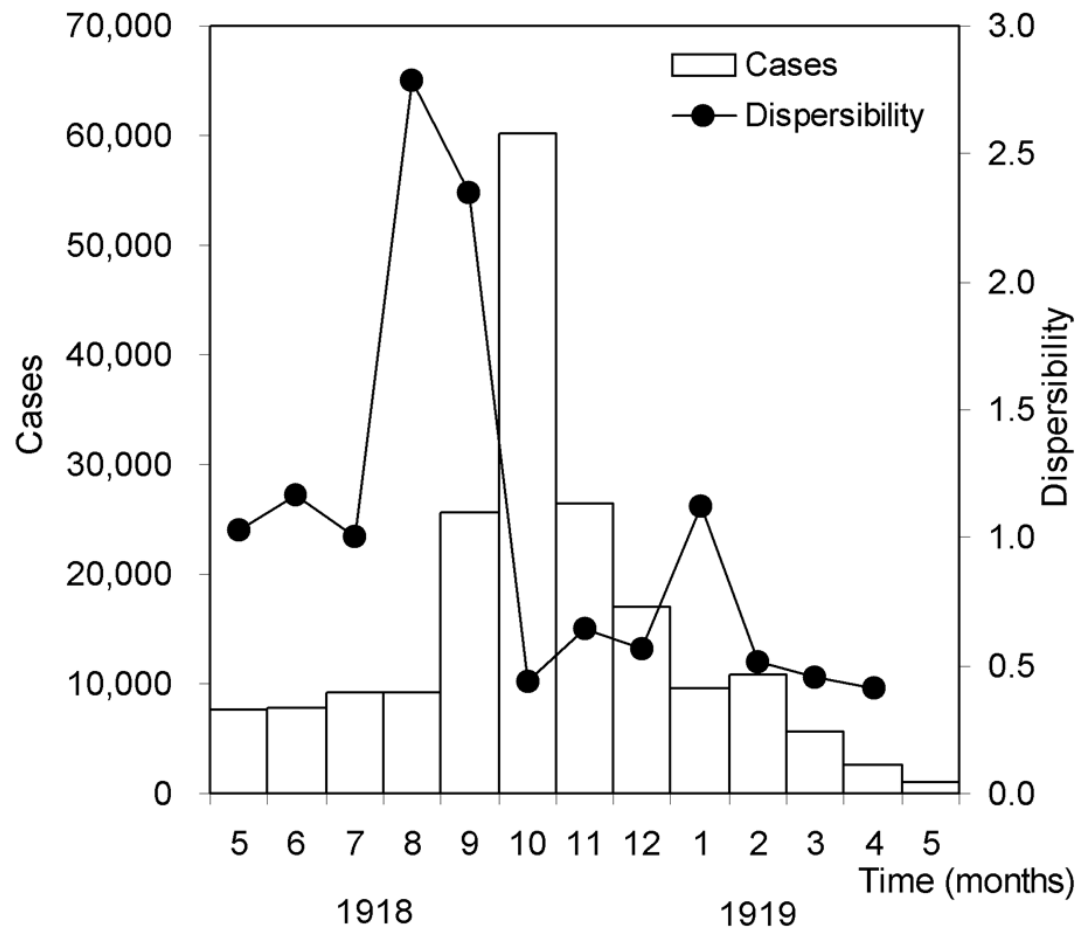
Clare Oswald Stallybrass (1881-1951)

defined as a marker that measures the "*total effect of factors affecting the spread of any specific infection at a given time and place*" the factors of which he discussed include "*sometimes intrinsic but more often depending upon either external or secondary factors.*"

Stallybrass CO. The principles of epidemiology. 1931.
London: George Routledge and Sons Inc.

Nishiura H, et al. J R Soc Interface 2010;7:297-307

the original (uncorrected) ratio of dispersibility following Stallybrass's exercise using monthly reports of cases due to pandemic influenza from 1918-19 in the French army (Delater, 1923)



Stallybrass's incorrect "correcting factor"

pp. 503

alternative method which would give only an approximation to accuracy would be the application of a correcting factor

$$= \frac{\text{length of successive periods in days}}{\text{mean incubation period in days}} \dots\dots\dots (ii)$$

Thus if it is desired to compare the dispersibilities of measles and influenza, and the incubation periods are assumed to be 11 days and 2 days, the correcting factor for statistics collected in weekly periods would be

$$(\text{measles}) \frac{7 \text{ days}}{11 \text{ days}} = 0.63, \text{ and } (\text{influenza}) \frac{7 \text{ days}}{2 \text{ days}} = 3.5.$$

pp. 504

~~A corrected ratio of dispersibility is obtained thus :~~

$$\frac{\text{cases in period B}}{\text{cases in period A}} \times \text{correcting factor} \dots\dots\dots (iii)$$

幾何級数的増殖モデルが教えてくれること

If the reporting interval, Δt , is exactly a multiple of the mean generation time, we obtain


$$J_k = a_0 + a_0 R + a_0 R^2 + \cdots + a_0 R^{n-1}$$

$$= a_0 \sum_{i=0}^{n-1} R^i$$

$$J_{k+1} = a_0 R^n + a_0 R^{n+1} + a_0 R^{n+2} + \cdots + a_0 R^{2n-1}$$

$$= a_0 \sum_{i=n}^{2n-1} R^i = a_0 R^n \sum_{i=0}^{n-1} R^i$$

$$n = \frac{\Delta t}{\mu} = \frac{\text{length of reporting interval}}{\text{mean generation time}}$$


$$\hat{R} = \left(\frac{J_{k+1}}{J_k} \right)^{\frac{1}{n}}$$

例えば

(直近7日間の患者数) < (その前7日間の患者数)

(直近7日間の患者数)/(その前7日間の患者数) < 1

は実効再生産数とコンセプトとして近い

また、目的が流行動態の理解のためなら、リアルタイム性を重視して、これを報告日ベースで計算しても構わない

Pitfall：精密な情報を失う、報告の遅れが問題

$R(t)$ の実践性のまとめ

- ・ 非定常状態での流行増減の記述
- ・ R_0 と一定の関係がある
- ・ 「増えているか否か」「どのスピードで増えているか」が記述可能
- ・ Non-linear phaseでの推定なのに、 $S(t)$ の情報なく流行曲線だけから推定できる。（ただし、 $S(t)$ が減ったのか対策が奏功したのか、など切り離せないことあり）