

2. 新型コロナウイルス感染症研究のデジタルトランス フォーメーション

朴 炯基¹⁾, 禹 周賢¹⁾, 岩波 翔也¹⁾, 岩見 真吾^{1,2,3,4,5)}

1) 名古屋大学大学院 理学研究科 理学専攻 生命理学講座 異分野融合生物学研究室

2) 九州大学マス・フォア・インダストリ研究所

3) 京都大学高等研究院 ヒト生物学高等研究拠点 (ASHBi)

4) 公益財団法人がん研究会 ネクストがん研プロジェクト

5) 理化学研究所数理創造プログラム

生命医科学研究領域では、さまざまな解析手法を駆使したビッグデータ利活用による新しい研究分野が台頭する時代になり、機械学習を中心としたデータ駆動型アプローチが注目を集めている。しかし、例えば、社会が直面している感染症・がん・臓器疾患に目を向けた場合、臨床試験や医療機関から得られる高精度（経時的）医療データには制約があり、その大量取得は容易ではない。特に、新興感染症発生時であれば、限られた症例数から得られる臨床データを用いて、治療戦略を開発したり、公衆衛生上の対策を策定する必要がある。つまり、多くの臨床データはビッグデータではないため、データ駆動型アプローチの適用がしばしば困難になる。本稿では、COVID-19の臨床データに対して、主にモデル駆動型アプローチを適用し、限られたデータから生物学的に重要な情報がどのように抽出され、また、それらがどのように社会に役立つのかを議論する。

はじめに

COVID-19は飛沫感染と接触感染が主体であり、潜伏期は1～14日、平均5～6日で発熱、咳嗽、倦怠感などの感冒症状がみられる。およそ80%の症例はそのまま改善するが、20%は肺炎を発症し入院や酸素投与が必要となり、5%程度は重症化して集中治療を要する。特に高齢、高血圧、糖尿病、肥満などが重症化のリスク因子である。また、無症状病原体保有者も数多く存在することが、感染拡大の一因となっている。COVID-19は唾液や鼻腔ぬぐい液からSARS-CoV-2の核酸を検出するPCR法やLAMP法で診断する。ウイルスの抗原を検知する抗原検査は、安価、簡便、

迅速であるが、現時点ではPCR法に比べて感度が低く、検査精度の改善が続けられている。

COVID-19の治療法確立を目指した臨床研究では、PCR検査によってウイルス量が検出できなくなるまでの期間（ウイルス排出期間）や一定期間における累積ウイルス量などが評価項目として用いられている。しかし、患者体内でSARS-CoV-2はどのように増加し、その後、どのように減少するかの定量的な知見は少ない。治療法の確立を加速させるためには、時々刻々と変化するウイルス感染動態を特徴づけることが重要である¹⁾。また、これらの情報は患者のウイルス量と症状の度合、重症化・致死率、伝播率などとの関係性を明らかにする上でも大いに役立つ²⁾。過去には、HIVやHBV、HCVなどの感染動態の理解が治療法の開発を後押ししてきた³⁾。なお、COVID-19に関連した症状を呈する人、感染者と濃厚接触した人、感染地域への渡航歴がある人に対して、感染症の診断のためにウイルス検査が推奨されている。よって、感染疑い発生時の診断目的でのウイルス検査だけでなく、無症候感染者のスクリーニングや有病率・感染動向の調査をする目的でもウイルス検査は重要である。

本稿では、ウイルス感染動態を分析する数理学的手法

連絡先

〒464-8602

愛知県名古屋市千種区不老町

TEL: 052-789-2992

FAX: 052-789-3054

E-mail: iwami.iblab@bio.nagoya-u.ac.jp

<https://iblab.bio.nagoya-u.ac.jp>

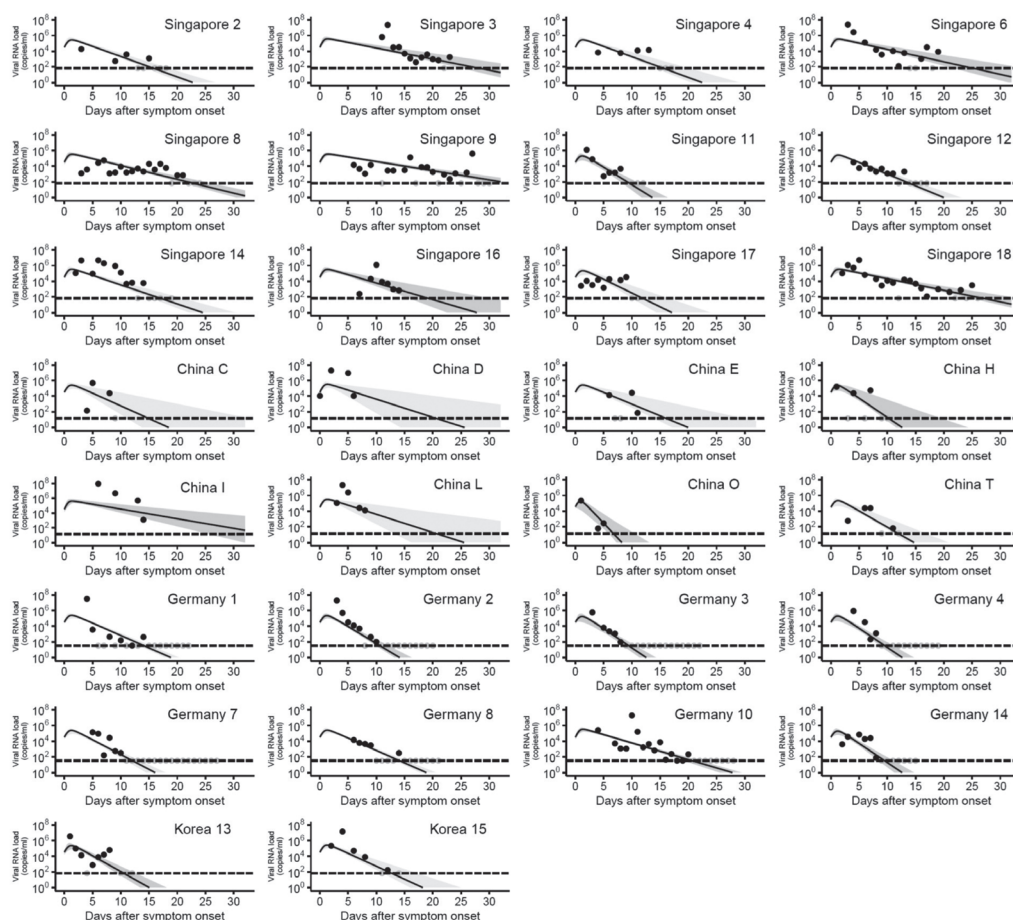


図1 COVID-19 感染者のウイルス感染動態：ウイルス量は入院した SARS-CoV-2 感染患者の鼻腔 nasal（中国：China），咽頭 pharyngeal（ドイツ：Germany），鼻咽頭 nasopharyngeal（シンガポール：Singapore，韓国：Korea）および中咽頭 oropharyngeal（韓国：Korea）のぬぐい液から計測された。PCR 法による SARS-CoV-2 の検出限界値は、点線で示した。図中の点と実線は、各患者の計測されたウイルス量と、個別のパラメータを用いて推定されたウイルス量である。塗りつぶされた領域は 95% 予測区間である。

を紹介し、ウイルス感染がどのように特徴付けられるのかを単純な数理モデルを用いて解説する。そして、数理モデルを用いて、新型コロナウイルスのウイルス排出期間を定量し、抗ウイルス薬による治療効果を評価する。最後に、ウイルス排出期間に基づいた網羅的シミュレーションを用いて感染者隔離戦略を評価する。これら数理科学を起点としたウイルス感染動態の解析から得た知見は、治療戦略の開発や感染症の防疫に役立ち、公衆衛生上の対策を策定する上で重要な役割を果たす。以下、私たちが取り組んできた新型コロナウイルス感染症研究のデジタルトランスフォーメーション（DX）に資する話題を簡単に紹介する。

コロナウイルス感染者における時系列ウイルス量の臨床データ

COVID-19 の流行発生当初から、各国で患者体内のウイルス量の時間変化が報告されている。これらの報告では、

スワブと呼ばれる綿棒のようなものを使って、鼻咽頭等の呼吸器から検体を採取し、PCR によって定量化されたウイルス量を計測している。COVID-19 が疑われる何らかの症状（咳や熱など）の発症日を基準とし、定量化されたウイルス量の値、もしくは PCR 陰性か陽性かどうかの質的情報がまとめられている。

本稿では、中国、韓国、シンガポール、ドイツの COVID-19 症例の中で、抗ウイルス治療が施されていない患者の時系列ウイルス量を解析した⁴⁸⁾。全ての症例でウイルス量は最終的に減少している一方で、その排出期間の長さは患者間で非常にばらつきが大きいことがわかる（図1）。

コロナウイルス感染者における時系列ウイルス量の定量的データ解析

一般にウイルスは、自身で増加することができず、宿主となる細胞内で宿主機能に助けられながら増殖する。ウイ

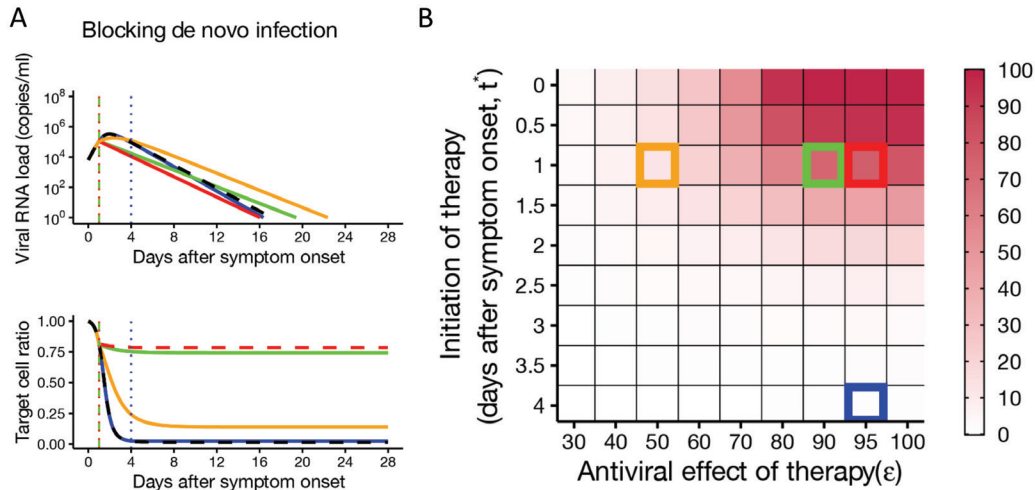


図2 COVID-19 感染者に対する薬剤治療効果：A. 抗体医薬や侵入阻害薬による新規感染阻害の治療効果が与えるウイルス感染動態への影響を計算した（上：ウイルス量 下：標的細胞の割合）. B. 抗ウイルス効果 ϵ と治療開始時刻 t^* の違いによる累積ウイルス量の減少率を計算した。図Aの縦点線および色は図Bの治療の種類に対応している。

ルスは細胞内へ侵入するために、ウイルス表面のウイルスタンパク質と標的となる細胞の表面に発現している受容体となるタンパク質との特異的な相互作用を利用する。新型コロナウイルスでは、細胞への侵入に必要な受容体の一つとして、アンジオテンシンインペルターゼ2 (ACE2) を利用することがわかっている。細胞内に侵入したウイルスのDNAやRNAからは、宿主細胞の機能によって、その情報に従ったウイルスタンパク質が合成されるとともに、遺伝情報の複製となる新たなDNAやRNAが作られる。複製されたDNAやRNAの遺伝情報は、ウイルスタンパク質と会合し、子孫ウイルス粒子が組み立てられ、細胞外へ放出される。そして、新たな標的細胞への侵入、自身の複製を繰り返すことで体内での感染が広がる。つまり、宿主となる生物個体への感染から、生体内では、宿主細胞への侵入、細胞内での複製と細胞外への子孫ウイルスの放出という一連のウイルス複製の流れが同時に繰り返されることで、ウイルス量が時間とともに変動していく。

生体内でのウイルス量の時間変化を記述する上で、筆者たちが開発した最も単純な数理モデルを紹介する。発症時 ($t=0$) の標的細胞数を基準とした時の時刻 t の標的細胞の割合を $f(t)$ 、ウイルス粒子数を $V(t)$ (copies/ml) と定義すると、上記のウイルス感染動態（ウイルスダイナミクス）は、以下の常微分方程式系であらわされる^{1,9,13)}；

$$\begin{aligned} \frac{df(t)}{dt} &= -\beta f(t)V(t), \\ \frac{dV(t)}{dt} &= \gamma f(t)V(t) - \delta V(t). \end{aligned} \quad (1)$$

ウイルス粒子の標的細胞への感染力は、ウイルス粒子が標的細胞と接触する頻度や、細胞表面への付着、細胞内への

侵入の効率などを反映した感染定数 β とウイルス粒子数の積とする。ウイルスが感染した細胞は、ウイルス自身や免疫応答の細胞障害性等により δ という率で排除される。また、 γ はウイルスの最大複製率とする。

数理モデル (1) では、標的となる細胞の外部からの流入を考慮せず、標的細胞数には限りがあると仮定している。また、COVID-19やインフルエンザなどの急性感染症は、比較的短期間でウイルスが増殖した後体内から除去され、その後持続的な感染が起こらないので、標的細胞の細胞死を無視している。

急性感染症患者から取得されるデータは、多くの場合、発症時以降の時系列ウイルス量である。数理モデル (1) では、標的細胞数の時間変化を発症時の標的細胞数で正規化しているので $f(0) = 1$ である。よって、データフィッティングから推定すべき未知パラメータは β 、 γ 、 δ 、 $V(0)$ の4つとなる。

コロナウイルス感染者から得られる時系列ウイルス量は、患者間で大きなばらつきが見られる。患者間の免疫応答の不均一性など様々な理由が考えられるが、それらは数理モデルの“パラメータの違い”に反映されると仮定できる。個人ごとのデータに対してパラメータ推定が実施できたら理想的であるが、症例ごとに検査する頻度や回数は異なり、また、個々人から得られるデータ数も限られるので、信頼できるパラメータ推定が実施できない症例もある。特に、パンデミック感染症の場合、人命救助が最優先事項になり、必然的に取得される検体やデータの質が下がるのが常である。

しかし、このように症例ごとのパラメータ推定に信頼性がない場合であっても、“症例群”としてのパラメータ推

定を実施する方法がある。それは非線形混合効果モデルと呼ばれる方法で、この数理モデルを使えば患者間で異なる効果（変量効果）と患者集団に共通する効果（固定効果）が推定できる。変量効果と固定効果は全患者の時系列データから推定されるので、データ数が少ない個人に対しても、他の患者が持つデータを踏まえた推定パラメータが割り当てられることになる。

非線形混合効果モデルのアプローチによるパラメータ推定の実装方法として、Monolix というソフトウェアがある (<http://lixoft.com/products/monolix/>)。Monolix では、それぞれのパラメータの個体ごとの値は $h(\theta_i) = h(\theta_{pop}) + \pi_i$ と表される。ここでは、固定効果 θ_{pop} と、平均 0、分散 Ω の正規分布に従う変量効果 $\pi_i (\sim N(0, \Omega))$ を使っている。また、関数 h は、個体ごとのパラメータの値が従う確率密度分布を決める関数で、Monolix (MonolixSuite 2020R1) では、正規分布、対数正規分布、ロジット正規分布、プロビット正規分布を用いることができる（本稿では対数正規分布を用いた）。そして、確率的近似 EM アルゴリズムによって、固定効果 θ_{pop} と変量効果の分散 Ω を推定できる。

推定したパラメータの固定効果を用いて数理モデル (1) から予測されるウイルス感染動態を実線で、また、症例ごとの予想から計算した 95% の予測区間を帯で示した (図 1)。

抗ウイルス薬剤による治療効果分析の DX

細胞から上気道（鼻・咽頭・喉頭）へ排出されるウイルス量がピークを迎える前に抗ウイルス薬剤の投与を開始することが、体内のウイルス排出量を減少させるために重要である。このことは、インフルエンザなどの臨床試験からも一般的に知られている¹⁴⁾。

ここでは、薬剤治療の効果を詳細に分析するために、数理モデル (1) を拡張し、ウイルスの感染阻害の影響を組み込むことで、治療下でのウイルス量の変化を評価する。感染阻害率 $\varepsilon (0 \leq \varepsilon \leq 1 : \varepsilon = 1$ のとき、ウイルス感染阻害の効果は 100% である) を用いれば、ウイルスの感染動態は以下の微分方程式で表される¹⁹⁻¹³⁾：

$$\begin{aligned} \frac{df(t)}{dt} &= -(1 - \varepsilon H(t))\beta f(t)V(t), \\ \frac{dV(t)}{dt} &= (1 - \varepsilon H(t))\gamma f(t)V(t) - \delta V(t). \end{aligned} \quad (2)$$

ここで、

$$H(t) = \begin{cases} 0 & t < t^* \\ 1 & t \geq t^* \end{cases}$$

であり、 t^* は治療を開始した時刻を表している。なお、抗体による中和作用やワクチン摂取による抗体価の上昇による新規感染の抑制なども同様に考えることができる¹⁾。数理モデル (2) と推定パラメータを用いて、阻害率 ε と治療開始時刻 t^* に対するウイルス量の変化が計算できるの

で、コンピュータシミュレーションによりさまざまな治療の効果を評価していく (図 2)。ここでは、阻害率が $0.3 \leq \varepsilon \leq 1.0$ の抗ウイルス薬剤を考え、また、発症後の治療開始時刻を $0 \leq t^* \leq 4$ とする (図 2A)。治療効果の指標は、累積ウイルス量 ($AUC : \int_0^{28} V(t) dt$) の減少率とした (図 2B)。なお、推定パラメータの固定効果を用いれば、非治療の場合、COVID-19 感染者の累積ウイルス量は 8.2×10^5 (RNA コピー数/ml) であり、ほとんど全ての標的細胞は枯渇していることが計算できる。

例えば、90% の阻害率をもつ薬剤治療が発症後 1 日で開始された場合、AUC は 73% まで減少し、また、74% もの標的細胞を守れることがわかった (図 2：緑)。他方、阻害率 95% の薬剤であっても治療が発症後 4 日も遅れて開始された場合、AUC は 14% 程度しか減少せずに、また、98% の標的細胞が感染してしまった (図 2：青)。つまり、薬剤の阻害率が高く、治療開始時刻が早ければ、良い治療効果が期待される。特に、抗ウイルス薬剤を用いた治療では、早期の治療開始が極めて重要であることがわかる。これは、ウイルス排出量がピークに達する頃には、すでに多くの標的細胞に感染が広がっていることが原因の 1 つと考えられる。

感染者の隔離効果分析の DX

国ごとに異なるが、現在、感染者を固定された期間だけ隔離する方法である固定期間法が多く採用されている。例えば、アメリカ疾病管理予防センター (CDC) のガイドラインでは、軽症例および中等症例では、「発症後 10 日間経過している」「解熱剤なしで最後の発熱から 24 時間が経過している」「咳や息切れの症状が改善している」という 3 つの条件を満たした場合、感染者隔離を終了できる、としている (いた)。しかし、このような固定期間法は、重症度や症状の継続に関連する個々の感染者のウイルス量の不均一性を考慮していないため、感染性を持った状態の患者の隔離を終了してしまった場合には、2 次感染を完全には防ぐことができない場合もある。一方、すでに感染性を失っているにも関わらず無意味に長期間隔離してしまう場合もある。

ここでは、ウイルス感染動態の数理モデル (1) と推定したパラメータを用いて、患者間のばらつきを再現した 1,000 個の“期待ウイルス量”を準備し、固定期間法による感染者隔離について詳細に分析する。なお、固定期間法を評価するために“不完全な隔離終了確率”と“不要な隔離期間”を計算した。不完全な隔離終了確率とは、まだ感染性のある隔離者であるにもかかわらず、隔離が終了してしまう確率と定義した。また、不要な隔離期間は、すでに感染性を失っているにもかかわらず、過剰に隔離してしまう期間と定義した。具体的には、隔離期間と感染性を持つ期間の平均値の差で定義する。なお、不要な隔離期間が負の値をと

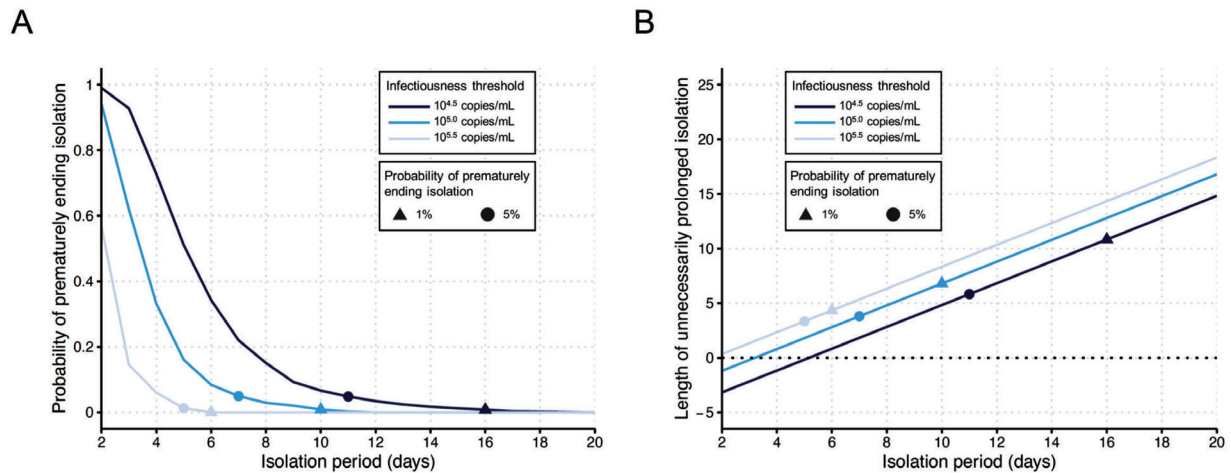


図3 固定期間法による隔離終了：A. 隔離期間に応じた不完全な隔離終了確率（まだ感染性のある隔離者であるにもかかわらず隔離が終了してしまう確率）を計算した。B. 隔離期間に応じた不要な隔離期間（すでに感染性を失っているにもかかわらず過剰に隔離してしまう期間）を計算した。▲と●はそれぞれ不完全な隔離終了確率が1%と5%に対応し、感染者が感染性を持つためのウイルス量閾値は $10^{4.5}/10^5/10^{5.5}$ の3つを考えている。

る場合は、隔離者が感染性を持つにもかかわらず隔離が終了してしまったこと、すなわち不完全な隔離終了を意味する。

まず、発症後固定した期間 (s_F と呼ぶ) だけ感染者を隔離すると仮定する。この場合、隔離終了時点において、期待ウイルス量が感染性閾値よりも大きくなっている確率、すなわち、不完全な隔離終了確率を p とする。この時、1,000 個の“期待ウイルス量”から計算される p の推定値 $\hat{p}$ は、 $\hat{p} = \sum_{i=1}^{1000} I(V_i(s_F) > IT) / 1000$ である。ここで、 $V_i(*)$ は症例 i の期待ウイルス量であり、 IT は感染者が感染性を持つ（2次感染を起こす）ためのウイルス量閾値である。 $V_i(s_F) > IT$ を満たせば $I(*) = 1$ 、満たさなければ $I(*) = 0$ とした。なお、二項分布を仮定すれば、 p の95%信頼区間は、 $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/1000}$ と計算できる。また、不要な隔離期間は、症例 i の期待ウイルス量が感染性閾値を下回った時刻を $\hat{s}_i$ とすれば、 $s_F - \hat{s}_i$ である。以下、10日間の隔離期間とした。なお、検査の開始日は、発症日とした¹²⁾。

発症後10日間の隔離期間を仮定した固定期間法では、不完全な隔離終了確率が0.9%程度であり（図3A）、不要な隔離期間は6.8日程度である（図3B）。また、 $IT=10^{4.5}$ と $10^{5.5}$ のウイルス量閾値の場合、不完全な隔離終了確率、不完全な隔離終了確率は6.6%と0.0%、不要な隔離期間は4.8日と8.3日になる。さらに、 $IT=10^{4.5}$ 、 10^5 、 $10^{5.5}$ の閾値に対して、5%以下の不完全な隔離終了確率を保証するためには、それぞれ、11日、7日、5日の隔離期間が必要になる。この場合、5.8日、3.8日、3.3日の不要な隔離期間がそれぞれ発生する。また、より低い不完全な隔離終了

確率（1%など）を保証するためには、隔離期間を長く設定すればよいが、それに伴い、大多数の感染者が不必要に長く隔離されるので、損失が生まれる。このように、社会的に許容可能なレベルでのける感染者隔離戦略をオンデマンドに計算することができる。

おわりに

いくつかのウイルス性急性呼吸器感染症は新興感染症として、21世紀以降に先進国で急速に広がりパンデミックを起こした。2003年のSARSや2012年のMERSの流行時、渦中にあった香港やシンガポール、韓国では、当時の経験が後の感染症対策に生かされ、2020年のSARS-CoV-2によるCOVID-19流行の早い段階で臨床データを収集し学術研究として報告してきた。流行の初期フェーズの短期間にほぼリアルタイムに蓄積される疫学・ゲノム・臨床データを、数理モデルやシミュレーション、インフォマティクスを総動員した汎用化プラットフォームにより分析し、病原体の疫学的、臨床学的、ウイルス学的特性を定量化することで、分析結果を医療や感染症対策の現場にフィードバックできる。このように、感染症流行の可能な限り早期の段階で、疾患の特性と緊急度に応じた標準治療や対処法を確立することを強力に支援するプラットフォームが準備できれば、流行の後期フェーズに訪れる医療キャパシティへの負のインパクトを最小限に留めることに貢献できるはずである。これらはすべて、数理科学との融合を果たすことで加速され、今後、このような異分野融合研究に基づいたデジタルトランスフォーメーションが期待されている。

謝辞

本特集執筆の機会をお与えいただきました野田岳志先生はじめ編集委員の先生方に感謝いたします。本稿に関連し、開示すべき利益相反状態にある企業等はありません。また、本稿で紹介した研究は、ムーンショット型研究開発事業（JPMJMS2021・JPMJMS2025）の一環で行われました。

参考文献

- 1) Kim, K.S., *et al.* A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLoS Biol* **19**, e3001128 (2021).
- 2) Marc, A., *et al.* Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *Elife* **10** (2021).
- 3) Perelson, A.S. Modelling viral and immune system dynamics. *Nat Rev Immunol* **2**, 28-36 (2002).
- 4) Kim, E.S., *et al.* Clinical Course and Outcomes of Patients with Severe Acute Respiratory Syndrome Coronavirus 2 Infection: a Preliminary Report of the First 28 Patients from the Korean Cohort Study on COVID-19. *J Korean Med Sci* **35**, e142 (2020).
- 5) Oh, M.D., *et al.* Viral Load Kinetics of MERS Coronavirus Infection. *N Engl J Med* **375**, 1303-1305 (2016).
- 6) Wölfel, R., *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465-469 (2020).
- 7) Young, B.E., *et al.* Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *Jama* **323**, 1488-1494 (2020).
- 8) Zou, L., *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* **382**, 1177-1179 (2020).
- 9) Ejima, K., *et al.* Time variation in the probability of failing to detect a case of polymerase chain reaction testing for SARS-CoV-2 as estimated from a viral dynamics model. *J R Soc Interface* **18**, 20200947 (2021).
- 10) Ejima, K., *et al.* Estimation of the incubation period of COVID-19 using viral load data. *Epidemics* **35**, 100454 (2021).
- 11) Iwanami, S., *et al.* Detection of significant antiviral drug effects on COVID-19 with reasonable sample sizes in randomized controlled trials: A modeling study. *PLoS Med* **18**, e1003660 (2021).
- 12) Jeong, Y.D., *et al.* Revisiting the guidelines for ending isolation for COVID-19 patients. *Elife* **10**(2021).
- 13) Kim, K.S., *et al.* Incomplete antiviral treatment may induce longer durations of viral shedding during SARS-CoV-2 infection. *Life Sci Alliance* **4**(2021).
- 14) Nicholson, K.G., *et al.* Efficacy and safety of oseltamivir in treatment of acute influenza: a randomised controlled trial. Neuraminidase Inhibitor Flu Treatment Investigator Group. *Lancet* **355**, 1845-1850 (2000).

Digital transformation of COVID-19 research

Hyeongki Park¹⁾, Joo Hyeon Woo¹⁾, Shoya IWANAMI¹⁾ and Shingo IWAMI^{1,2,3,4,5)}

1) interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University

2) Institute of Mathematics for Industry, Kyushu University

3) Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University

4) NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR)

5) Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN

In a current life sciences research, we are in an era in which advanced technology emerging and utilize big data. Data-driven approaches such as machine learnings play an important role to analyze these datasets. However, limited clinical (time-course) datasets are available for infectious diseases, cancer, and other diseases. Especially in the case of emerging infectious disease outbreaks, clinical data obtained from a limited number of cases must be used to develop treatment strategies and public health policies. This means that many clinical data are not big data, which often makes the application of data-driven approaches difficult. In this paper, we mainly apply a mathematical model-based approach to the clinical data of COVID-19 and discuss how biologically important information can be extracted from the limited data and how they can benefit society.

