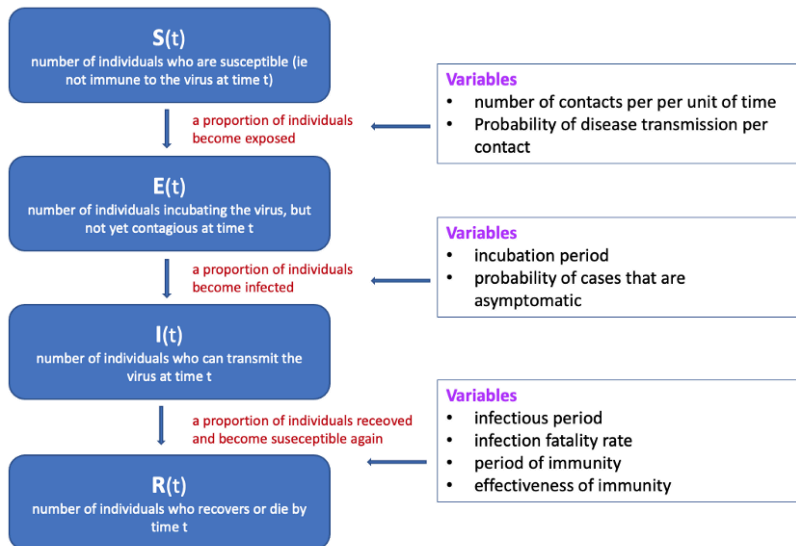


Gs. Nguyễn Văn Tuấn - Tất cả mô hình tiên lượng Covid đều sai!

24/8/2021



Minh hoạ cho mô hình SEIR. Nguồn:

<https://tuanvnguyen.medium.com/move-away-from-lockdowns-ad24fe2d995b>

Nhiều bạn thắc mắc tại sao mô hình dự báo covid ở VN quá sai. Tôi dành ra khá nhiều thì giờ để suy nghĩ và làm về tiên lượng [1], nên có thể chia sẻ những khó khăn đằng sau để các bạn biết. Không phải ai có kĩ năng toán hay thống kê học là có thể xây dựng mô hình tốt đâu. Xây dựng mô hình tiên lượng tốt đòi hỏi những kiến thức chuyên ngành, kinh nghiệm thực tế, và kĩ thuật thật tốt. Cái note này giải thích tại sao các mô hình covid đều sai và nên sử dụng mô hình như thế nào.

Có lẽ các bạn còn nhớ tuần qua tôi có ‘tâm sự’ nỗi khổ của người làm Academic Editor khi phụ trách mấy bài nghiên cứu về covid. Trong cái note đó tôi có đề cập đến Giáo sư Neil Ferguson (một chuyên gia lừng danh về dịch tễ toán học ở Anh và là cố vấn cho Chính phủ Anh về Covid) xây dựng một mô hình tiên lượng nổi tiếng trên thế giới. Mô hình này ước lượng rằng dịch Vũ Hán sẽ gây ra 500,000 cái chết nếu Chính phủ Anh không có hành động.

Chánh phủ Anh lập tức ra chánh sách hạn chế đi lại, giãn cách xã hội, và phong tỏa. Đó là một mô hình có ảnh hưởng lớn trên thế giới.

Thế nhưng mô hình với 15,000 dòng codes [2] của Ferguson sai. Sai nghiêm trọng. Những gì xảy ra không như mô hình của ông tiên lượng. Vài nhóm nghiên cứu trên thế giới chỉ ra một số sai sót về giả định và kĩ thuật của mô hình. Có người thậm chí còn chê rằng những mã / codes trong mô hình đó chưa đạt trình độ chuyên nghiệp [3]!

Ngay cả trước đây ông cũng có mô hình tiên lượng số ca tử vong liên quan đến dịch cúm gia cầm (bird flu) sẽ gây ra 200 triệu người chết trên thế giới. Thế nhưng trong thực tế thì chỉ có 282 người chết trong thời gian 2003 đến 2009. Mô hình này cũng sai nghiêm trọng, và có chuyên gia chỉ ra rằng ông đã dùng chương trình máy tính quá xưa để dự báo cho dịch ngày nay.

Rồi chính ông vi phạm qui định về phong tỏa năm 2020, và thế là Ferguson quyết định từ chức cố vấn cho Chánh phủ Anh. Một kết cục buồn. Nhưng tôi hoàn toàn hiểu và thông cảm cho nỗi khổ của ông ấy. Không ai hoàn hảo, huống chi là một mô hình với 15,000 dòng codes.

Ở bên Mỹ, một trong những mô hình được Chánh phủ đề cập nhiều là của Institute of Health Metrics and Evaluation (IHME) [4]. Mô hình này dự báo số ca tử vong cho mỗi tiểu bang trong 4 tháng. Nhìn bề ngoài thì nó có vẻ ok đối với người không am hiểu dịch tễ học, nhưng nhìn từ góc độ chuyên môn thì mô hình có vấn đề. Chẳng hạn như mô hình này giả định rằng hệ số phản ánh ảnh hưởng giãn cách xã hội giống như những gì quan sát ở Vũ Hán, trong khi môi trường sống ở Vũ Hán rất khác so với Mỹ.

Ngay cả con số tử vong dự báo của mô hình cũng chẳng dựa vào dịch tễ học mà lệ thuộc vào số ca tử vong trước đây, nhưng số ca tử vong đó được báo cáo không đầy đủ. Hậu quả là mô hình đầy triển vọng này cho ra nhiều dự báo sai nghiêm trọng, ảnh hưởng đến việc hoạch định chánh sách.

Chỉ vài ví dụ trên để chúng ta thấy rằng mô hình toán học và mô hình dịch tễ học chỉ là một dạng ‘thể thao trí não’ (không phải ‘trí tuệ’ nghen). Mô hình là những công cụ có ích, nhưng đừng nên quá lệ thuộc vào chúng để dự báo tương lai xa. Tôi sẽ giải thích dưới đây tại sao mô hình rất dễ sai. Và chính vì sai nên giới lãnh đạo không nên quá lệ thuộc vào mô hình để hoạch định chính sách có thể gây ảnh hưởng đến hàng triệu người.

1. Hai bộ môn dịch tễ học

Tiên lượng bệnh tật là chuyên môn của giới dịch tễ học. Nhưng có lẽ đa số các bạn, kể cả người trong ngành y, không biết rằng có hai nhánh dịch tễ học. Một nhánh tạm gọi là Dịch tễ học Lâm sàng (clinical epidemiology), là các môn học liên quan đến ứng dụng các khái niệm căn bản như thiết kế nghiên cứu, điều tra dịch tễ, diễn giải và ứng dụng kết quả cho điều trị, v.v. Đây là những môn hay được giảng dạy cho bác sĩ và nhân viên y tế nói chung. Cá nhân tôi và vài đồng nghiệp hay tổ chức những lớp học như thế này ở Việt Nam và vài nước Á châu.

Một nhánh khác gọi là Dịch tễ Toán học (mathematical epidemiology). Đây là môn học đòi hỏi người theo học phải biết về y khoa, bệnh truyền nhiễm, và giải toán thống kê và máy tính. Người theo học phải có kiến thức về những môn như xác suất, qui trình ngẫu nhiên, phương trình vi phân, phương pháp Bayes, mô phỏng, v.v. Rất ít ai dạy môn này ở đại học cho bác sĩ hay các nhà dịch tễ học lâm sàng, nhưng những ai làm nghiên cứu thì phải học môn này.

Có thể nói rằng 100% người Việt chúng ta chỉ học dịch tễ học lâm sàng. Còn dịch tễ toán học thì rất hiếm ai học qua một cách bài bản, hay có kinh nghiệm ở trình độ nghiên cứu (research level).

Khi nói về mô hình dịch covid, chúng ta nói về nhánh dịch tễ toán học, chứ không phải dịch tễ học lâm sàng. Lí do đơn giản là các phương pháp dịch tễ toán học cung cấp hàng loạt ‘phương tiện’ để mô hình hoá. Các mô hình này có thể chia thành 2 nhóm: tiên lượng và dự báo.

Tiên lượng và Dự báo

Có lẽ vài bạn chưa hiểu rõ sự khác biệt giữa 2 khái niệm prediction (tiên lượng) và forecasting (dự báo), nên tôi xin có một giải thích ngắn gọn như sau:

Prediction hay tiên lượng là dùng tham số để giải thích một biến cố trong tương lai. Ví dụ như tôi dùng các yếu tố như hệ số lây nhiễm và cơ cấu dân số để hiểu dịch đang xảy ra như thế nào và dùng đó để giải thích rằng ảnh hưởng của phong tỏa là như thế nào.

Forecasting hay dự báo là sử dụng dữ liệu quá khứ để ước tính xác suất một biến số xảy ra trong tương lai. Chẳng hạn như tôi dùng số ca nhiễm hàng ngày trong tháng qua để dự báo số ca nhiễm tháng tới với một xác suất nào đó.

Những mô hình tiên lượng và dự báo trong dịch tễ học thì rất nhiều. Các bạn có thể xem qua trang này [5] để biết trên thế giới người ta đã làm đến đây về mô hình dịch tễ học. Tôi nghĩ tựu trung lại có 3 dạng mô hình chính:

Mô hình SEIR

Mô hình Agent

Mô hình thống kê

Dưới đây, tôi xin giải thích cho các bạn rõ hơn sự khác biệt giữa 3 dạng mô hình đó. Nên nhớ rằng mỗi dạng như thế có hàng chục mô hình khác nhau (chớ không phải chỉ 1), tùy thuộc vào sự phức tạp của môi trường dịch tễ.

2. Mô hình SEIR (susceptible – exposed – infection – recovery)

Mô hình SEIR đại khái phân chia quần thể thành 4 tình trạng (còn gọi là compartment): S là số người có thể bị nhiễm; E là số người bị phơi nhiễm; I là số người có thể lây lan cho người khác; và R là số người bình phục.

Mỗi chuyển biến từ S sang E, từ E sang I, và I sang R chịu sự chi phối của hàng loạt tham số. Chẳng hạn như từ S sang E, nhà nghiên cứu phải biết số lần tiếp xúc trong mỗi đơn vị thời gian và xác suất chuyển từ S sang E. Tương tự, từ E sang I lệ thuộc vào thời gian ủ bệnh và xác suất những ca không triệu chứng. Cụ thể, mô hình này đòi hỏi nhà nghiên cứu phải có những dữ liệu sau đây:

Dân số phân bố theo từng độ tuổi và giới tính;

Xác suất số ca nhiễm nhẹ sẽ đến khám ở phòng mạch bác sĩ;

Xác suất cổ ca nhiễm nhẹ nhập viện;

Xác suất số ca do bác sĩ gia đình điều trị sẽ nhập viện;

Xác suất ca xuất viện và cần bác sĩ điều trị;

Xác suất số ca nhiễm nặng;

Xác suất số ca nhiễm nặng cần nhập viện;

Xác suất số ca nhiễm nặng cần nhập ICU;

Xác suất số ca nhiễm nặng cần thở máy;

Phân bố thời gian nằm viện;

Phân bố thời gian nằm ICU;

Chỉ số turnover giường bệnh và ICU.

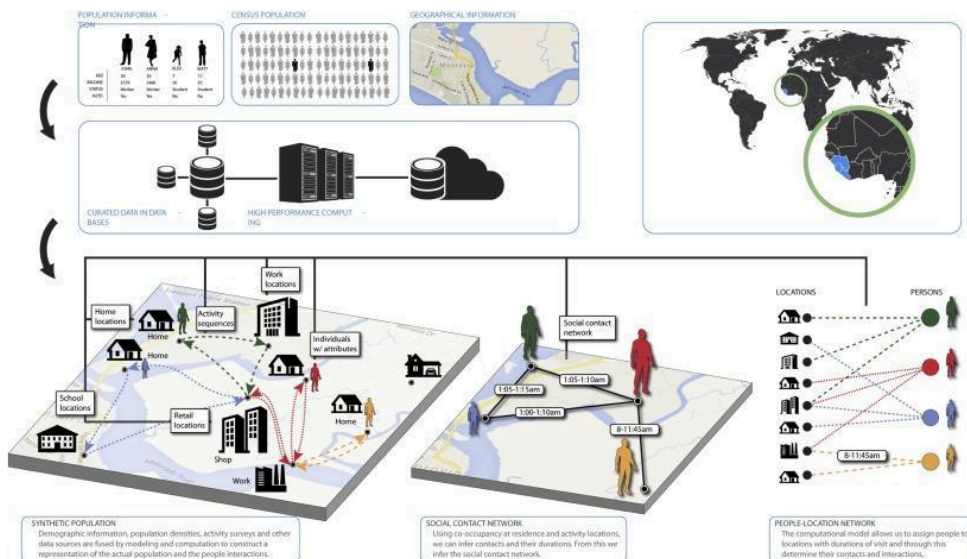
Tất cả những tham số đó rất khó có. Khi dịch mới khởi phát thì làm sao nhà nghiên cứu có những tham số đó. Thành ra, cách mà họ làm là đưa ra một con số nào đó cho mỗi tham số. Nhưng thử tưởng tượng, một mô hình như SEIR nếu đơn giản thì cần đến ít nhất là 9 tham số, tức là nhà nghiên cứu có hàng trăm mô hình khả dĩ.

Trong tình huống thiếu thốn dữ liệu thì chỉ có cách là mô phỏng. Nhưng mô phỏng lại tùy thuộc vào hàng loạt tham số kĩ thuật khác mà nhà nghiên cứu có khi không biết hết, và một lần nữa họ phải ... võ đoán.

3. Mô hình Agent

Mô hình này còn gọi là ABM (Agent-based Model) trong dịch tễ học. Mô hình ABM hoàn toàn dựa vào mô phỏng bằng máy tính loại công suất cao (như supercomputer chẳng hạn). ABM thường được ứng dụng để nghiên cứu về tương tác giữa người với người, giữa các môi trường, địa phương, và thời gian. Đây là mô hình hoàn toàn dựa vào xác suất, và có thể xem như là một dạng 'system modeling' chứ không đơn giản như mô hình thống kê truyền thống.

Các bạn thử tưởng tượng người ta mô hình các mối tương tác giữa 10 triệu người thuộc 100 địa điểm và suốt 365 ngày. Ông A gặp ông B ở địa điểm 1 trong thời gian t_1 ; ông Ba tiếp xúc bà C tại địa điểm 2 trong thời gian t_2 ; v.v. Rất phức tạp và tùy thuộc vào rất nhiều tham số. Vì tính phức tạp, nên mô hình này thường đòi hỏi phải có máy tính công suất lớn, kể cả siêu máy tính.



Minh hoạ cho mô hình ABM. Nguồn:

<https://www.sciencedirect.com/science/article/pii/S1755436517300221>

Nhưng mô hình ABM nếu đúng thì có thể giúp nhà nghiên cứu biết được các biện pháp giãn cách xã hội có hiệu quả như thế nào.

4. Mô hình thống kê truyền thống

Đây là các mô hình tôi xếp vào nhóm ‘truyền thống’ như mô hình dữ liệu dãy thời gian (time series models như ARMA, ARIMA, SARIMA), mô hình hồi qui (regression models), hay những kĩ thuật gọi chung là ‘curve fitting’. Những mô hình này nói chung là khá đơn giản, vì chỉ sử dụng con số ca nhiễm hay tử vong mỗi ngày và tìm một phương trình tiên lượng cho tương lai.

Các mô hình thống kê này chỉ có thể ứng dụng cho dịch trong một thời gian ngắn và đơn giản. Tuy vậy, các mô hình này vẫn phải dựa vào nhiều giả định liên quan đến kĩ thuật và can thiệp. Chẳng hạn như nhà nghiên cứu phải giả định rằng biện pháp can thiệp sẽ diễn ra vào lúc nào và hiệu quả ra sao, tức là những tham số không thể nào biết trước được.

Đối với dịch phức tạp và có nhiều can thiệp như covid thì những mô hình truyền thống này không được ứng dụng. Dĩ nhiên, trong thực tế cũng có người ‘mạo hiểm’ ứng dụng các mô hình này, nhưng dân trong chuyên ngành chỉ cần nhìn qua là biết không phải người trong ‘bộ lạc’.

Dĩ nhiên, còn nhiều mô hình khác nữa mà tôi không đề cập ở đây. Chẳng hạn như có người dùng Machine Learning để dự báo dịch covid cũng rất hay [6]. Nhưng cá nhân tôi thì không mặn mà với mô hình loại này, vì nó thiếu các nguyên lí dịch tễ đằng sau. Mô hình cần phải có lí thuyết và nguyên lí, chứ thiếu yếu tố này thì chỉ là ‘hộp đen’ mà thôi.

5. Mô hình nào cũng sai

Ở trên, tôi đã giải thích sự phức tạp và qui mô của mô hình tiên lượng SEIR và ABM. Phức tạp là vì các mô hình này tùy thuộc vào quá nhiều tham số và giả định. Những tham số đó thường không có sẵn. Còn giả định thì có thể sai. Sai là vì nhà nghiên cứu nhìn hiện tượng theo cảm nhận của mình, chứ ít khi nào theo sự vật. Nói như Einstein là “assumptions are

made and most assumptions are wrong” (Giả định là cái mình tự đặt ra và đa số giả định đều sai).

Như tôi trình bày trên, tính hợp lí của mô hình lệ thuộc vào 3 yếu tố: giả định và phẩm chất của dữ liệu. Một trong những giả định quan trọng là thời gian ủ bệnh, hệ số ảnh hưởng của giãn cách xã hội, và chỉ số lây lan (reproduction ratio). Vấn đề là trong thời điểm dịch mới bùng phát thì dữ liệu rất hạn chế, có thể do báo cáo tư cơ sở không đầy đủ, do xét nghiệm kém chính xác, hay do số liệu được thu thập còn quá rời rạc. Một trong những vấn đề về dùng dữ liệu gộp từ nhiều nguồn là nó bỏ qua sự khác biệt giữa các địa phương. Những yếu tố lâm sàng như bệnh đi kèm, độ tuổi, yếu tố nguy cơ, v.v. bị trung bình hoá, thành ra dẫn đến một nghịch lí là mô hình có vẻ rất tốt cho tất cả địa phương, nhưng lại rất kém cho mỗi địa phương!

Các mô hình phải có tính minh bạch. Minh bạch hiểu theo nghĩa nhóm nghiên cứu nên công bố chi tiết kĩ thuật về mô hình cho thế giới xem và đánh giá. Điều này cũng giống như làm nghiên cứu khoa học thì phải công bố kết quả và phương pháp nghiên cứu. Chẳng hạn như nhóm của Neil Ferguson [7] hay nhóm Doherty [8] người ta công bố mô hình bằng những bài báo khoa học hay trên các trạm như MedrXiv.

Nhà thống kê học trứ danh George Box [9] (con rể của Ronald Fisher) từng nói một câu nổi tiếng rằng ‘Tất cả các mô hình đều sai, nhưng một số mô hình vẫn có ích’ (‘All Models Are Wrong, Some are Useful’). Nhiều người hiểu sai câu này và biện minh cho cái sai của mình. Nhưng câu đó phải hiểu đúng như sau: tất cả các mô hình đều sai, bởi vì bản chất của mô hình là xấp xỉ với thực tế (và xấp xỉ thì phải sai).

Tại sao ‘một số mô hình có ích’? Tại vì các mô hình được xây dựng theo đúng phương pháp và có dữ liệu tốt thì nó vẫn có thể giúp cho việc hoạch định chính sách và nhìn thấy trước viễn cảnh. Câu nói đó không biện minh cho những mô hình sai từ kiến thức đến phương pháp.

6. Tiêu chuẩn 4R cho mô hình tiên lượng

Để xây dựng một mô hình nghiêm chỉnh, một mô hình mà tôi gọi là ‘actionable model’ (tức mô hình có thể giúp lãnh đạo hành động) thì đòi hỏi nhiều hơn là một bài tập toán dịch tễ học. Bây giờ nghĩ lại, tôi thấy giá trị của một mô hình actionable được xác định bởi người làm ra nó và 4 tiêu chuẩn mà tôi gọi là 4R.

Ai cũng có thể xây dựng một mô hình tiên lượng hay dự báo, nhưng mô hình đó có được ‘công nhận’ hay không và công nhận bởi ai là một chuyện khác. Mô hình tiên lượng như là một công cụ, và công cụ có thể xem như một món hàng. Khách hàng khi lựa chọn mua món hàng, họ muốn biết ai là người sản xuất ra món hàng, nhà sản xuất đã có kinh nghiệm làm món hàng như thế chưa, và uy tín của nhà sản xuất ra sao. Một nhà sản xuất dùng một cái tuyên bố rằng họ làm ra [ví dụ] thuốc điều trị ung thư, nhưng họ chưa bao giờ có tên tuổi trong ngành, thì rất khó được công nhận. Trong mô hình tiên lượng cũng vậy, tác giả mô hình là một yếu tố khá quan trọng.

Xin chia sẻ một câu chuyện trong chuyên ngành loãng xương chúng tôi. Chuyên ngành này có 3 mô hình tiên lượng (một của nhóm tôi, một của WHO, và một của một nhóm bên Anh), nhưng chỉ có mô hình của nhóm tôi và WHO là được đồng nghiệp công nhận và sử dụng trong lâm sàng. Tại sao mô hình bên Anh cũng làm khá tốt mà không được sử dụng? Phải một thời gian sau người ta mới nhận ra lí do: vì họ không phải là người trong ‘bộ lạc’. Họ không có những ‘track record’ như nhóm chúng tôi và WHO (hiểu theo nghĩa từng công bố những nghiên cứu trong lãnh vực mô hình tiên lượng và quá quen mặt trong các hội nghị quốc tế). Họ xuất hiện khá ... đột ngột. Một lí do khác là mô hình của họ quá phức tạp để bác sĩ có thể sử dụng.

Mô hình tiên lượng covid cũng vậy. Trong hàng trăm mô hình khắp thế giới, người ta biết chọn mô hình nào? Đầu tiên là người ta phải xem thành tích khoa học của tác giả, họ xuất phát từ đâu và đã từng làm trong lãnh vực này bao lâu. Chẳng hạn như ở Úc này, giữa mô hình của nhóm Viện Doherty và Viện Burnet, có lẽ đa số sẽ chọn mô hình của Viện Doherty vì người đứng đầu có một thành tích khoa học khá lâu trong lãnh vực này.

Muốn xây dựng một mô hình tiên lượng tốt, cần phải đạt tiêu chuẩn gì? Xin ‘khoe’ với các bạn là cách đây 5 năm khi tôi được bổ nhiệm ở UTS [1], và theo truyền thống của đại học, tôi có một bài nói chuyện gọi là ‘ra mắt’. Bài này khá quan trọng vì qua đó mà người ta đánh giá mình trong faculty của đại học. Trong bài nói chuyện đó, tôi nói về “Predictive Medicine in the Genomic Era”, và đề ra tiêu chuẩn 4R cho một mô hình lí tưởng. Tôi có đề cập đến 4 tiêu chuẩn này trong cuốn sách mới xuất bản ở Việt Nam [10].

Reliable (tin cậy)

Relevant (liên đới)

Real world (thực tế)

Realtime (trực tuyến)

Tiêu chuẩn Reliable ở đây có nghĩa là mô hình phải đạt độ chính xác có thể chấp nhận được (nhưng không thể 100%). Để đạt được độ chính xác đó thì giá trị tiên lượng của mô hình phải unbiased, có nghĩa là không chịu ảnh hưởng bởi các yếu tố nhiễu. Ngoài ra, mô hình đó phải mang tính tái lập (reproducible), có nghĩa là người ta có thể lặp lại mô hình — nếu họ muốn.

Bất cứ mô hình nào cũng phải qua giai đoạn xây dựng trước rồi kiểm định sau. Có khi mô hình xây dựng rất tốt trên nền một dữ liệu ở Anh, nhưng có thể khi kiểm định ở Việt Nam thì sẽ thất bại. Chuyện này xảy ra khá thường xuyên. Tuyên bố rằng mình đã có một mô hình vẫn chưa đủ, mà phải chứng minh rằng mô hình đó reliable thì mới đáng xem xét (chưa nói đến sử dụng).

Tiêu chuẩn Relevance đề cập đến mô hình phải ‘actionable’, tức có thể dựa vào đó mà hành động. Nói cách khác, mô hình có thể dùng làm cơ sở để hoạch định chính sách. Nhiều mô hình xây dựng cho đẹp, nhưng không sử dụng được vì có thể nó chỉ là ‘hộp đen’.

Có những mô hình có quá nhiều tham số và giả định, mà người sử dụng rất khó kiểm tra tính hợp lí của chúng. Những mô hình này có thể cho ra kết quả tiên lượng hay dự báo khá tốt,

nhưng nó không mang tính liên đới. Chẳng hạn như mô hình với 20 tham số, và có tham số liên quan đến yếu tố A, và người sử dụng mô hình sẽ hỏi “Nếu tôi can thiệp vào A, thì kết quả sẽ ra sao?” Nếu kết quả can thiệp A không làm thay đổi giá trị tiên lượng thì mô hình đó được xem là irrelevant, không mang tính liên đới. Một mô hình loại đó chỉ làm như là một trò toán học thôi, không có giá trị thực tế.

Tiêu chuẩn Realworld là phải phản ánh thực tế và có giá trị kinh tế. Đó là mô hình phản ánh thực trạng của một địa phương, thay vì mô hình chung chung. Mỗi địa phương có những yếu tố đặc thù, và mô hình lí tưởng cần phải xem xét đến các yếu tố đó.

Có nhiều mô hình khả dĩ để xây dựng, và người xây dựng phải dung hoà giữa 2 thái cực: phức tạp và đơn giản. Mô hình đơn giản thì không phản ánh đầy đủ sự phức tạp của dịch. Còn mô hình quá phức tạp thì khó diễn giải, và có khi không phù hợp với dịch tễ và sinh học. Do đó, phải tìm một mô hình trung dung giữa 2 thái cực để phù hợp với tình hình thực tế.

Tiêu chuẩn Realtime là phản ánh sự biến chuyển theo thời gian của các yếu tố nguy cơ hay các tham số. Nói theo triết lí Vô Thường là không có yếu tố nào là cố định cả, cho nên mô hình tốt phải uyển chuyển theo thời gian. Đa số các mô hình tiên lượng thất bại ở tiêu chuẩn này vì giả định rằng các tham số của mô hình là cố định.

Trong các mô hình tiên lượng hay dự báo covid, cái sai sót lớn nhất là thiếu tính realtime. Chẳng hạn như quay lại các giả định trong mô hình SEIR, chúng ta thấy tất cả đều có thể thay đổi theo thời gian. Vả lại, nhà chức trách sẽ áp dụng các biện pháp giãn cách xã hội khác để kiểm soát dịch, nhưng nhiều mô hình tiên lượng không xem xét đến — hay có xem xét nhưng không đầy đủ — dễ dẫn đến sai lầm trong tiên lượng.

7. Hướng đi tương lai: hợp tác

Khi đại dịch xảy ra, có nhiều cá nhân và nhóm tự hình thành để góp một vài ý kiến hay việc làm cụ thể. Điều đó cũng dễ hiểu, vì mỗi chúng ta đều có tấm lòng với quê hương. Nhưng sự bất lợi của các nhóm tự phát là rời rạc và không có liên kết chặt chẽ. Một điểm yếu khác có

lẽ là sự tự tin của các cá nhân rằng họ là chuyên gia và biết tất cả, mà không cần tham vấn những người ngoài nhóm.

Nói chung, sự hình thành các nhóm ‘tự phát’ phản ánh một phần cái văn hoá của người Việt chúng ta. Nhiều người trong chúng ta hay nghĩ mình là nhút, và vì thế họ thấy không cần hợp tác với người ngoài có chuyên môn nhưng họ bất đồng chánh kiến.

Trong khi ở Úc và các nước phương Tây thì người ta có văn hoá hợp tác rất tốt. Chẳng hạn như các chuyên gia ở Mỹ và Úc, họ qui tụ thành các nhóm và làm tư vấn cho chánh phủ. Có nhóm chuyên về lâm sàng và y học thực chứng, chuyên đánh giá chứng cứ khoa học của các nghiên cứu và ra khuyến cáo. Có nhóm chuyên về tiên lượng và dự báo, huy động những chuyên gia từ các bộ môn dịch tễ học lâm sàng, dịch tễ toán học, toán học, thống kê học, khoa học máy tính, IT, vật lí, v.v. Kết quả là họ có những mô hình khả dĩ dùng làm cơ sở khuyến cáo chánh phủ.

Tôi nghĩ hướng đi tương lai là chúng ta nên hợp tác với nhau và làm việc vì mục tiêu chung. Chúng ta phải học cách làm việc của người phương Tây, vì sự phức tạp của các mô hình tiên lượng đòi hỏi người ta phải làm việc theo nhóm. Trong nhóm phải có người ‘chỉ huy’ cấp giáo sư có ‘track record’ trong chuyên ngành, và có nhiều chuyên gia có kinh nghiệm thực tế tốt. Các bạn thử xem cách Viện Doherty của Úc xây dựng mô hình tiên lượng ra sao. Ai cũng có thể xây dựng mô hình, nhưng mô hình có sử dụng hay có ích hay không còn tùy thuộc vào uy tín khoa học của người lãnh đạo.

Tôi hình dung ra nhóm nghiên cứu có những ‘tổ’ nhỏ như sau. Tổ chuyên thu thập và đánh giá phẩm chất dữ liệu. Tổ phân tích dữ liệu dịch tễ học và công bố báo cáo thường xuyên cho công chúng biết sự diễn biến của dịch. Tổ chuyên nghiên cứu mô hình dịch tễ học thích hợp nhất cho Việt Nam; tổ này bao gồm những người có chuyên môn về dịch tễ toán học. Tổ chuyên viết chương trình máy tính và xây dựng website, hỗ trợ tổ phân tích và tổ mô hình. Mỗi tổ có một người đứng đầu và chịu trách nhiệm với trưởng nhóm.

Tóm lại, một mô hình tiên lượng dựa vào rất nhiều giả định và có nhiều tham số mà nhà nghiên cứu không bao giờ có đầy đủ. Ở Việt Nam tình trạng thiếu thốn dữ liệu còn trầm trọng

hơn nữa, mà ngay cả có dữ liệu thì chất lượng rất kém. Xây dựng mô hình trên dữ liệu với chất lượng thấp thì kết quả sai là đương nhiên. (Cũng xin nhắc các bạn đừng gửi msg cho tôi nói rằng Việt Nam có đầy đủ dữ liệu. Mất thì giờ. Không có đâu. Loại dữ liệu cần cho mô hình không phải là dữ liệu hành chánh). Tôi nghĩ hướng đi, nếu ai còn tha thiết, cho tương lai có lẽ là:

(1) Hợp tác. xây dựng một mô hình tiên lượng là một việc làm rất khó khăn, đòi hỏi sự đóng góp của nhiều người có kinh nghiệm cao, và cần thời gian để thử nghiệm trước khi triển khai. Không thể và không nên xây dựng những mô hình theo ý mình (đa số giới khoa học là vậy) mà nên mô hình theo hiện tượng thực tế.

(2) Dữ liệu. Mô hình nào cũng cần dữ liệu với phẩm chất cao, bởi vì dữ liệu hay đầu vào là rác thì đầu ra của mô hình cũng là rác rưởi mà thôi (Garbage in, Garbage out). Do đó, cần phải có hệ thống thu thập dữ liệu cho thật tốt để phục vụ cho phân tích, và điều này thì tôi đã có đề nghị cụ thể trước đây [8].

(3) Mô hình chỉ là tham khảo. Tất cả các mô hình tiên lượng và dự báo không bao giờ được xem là ‘sự thật’ hay ‘fact’, bởi vì các mô hình này dựa vào nhiều giả định và phán xét của tác giả. Do đó, lệ thuộc vào mô hình để ra chánh sách là một sai lầm nguy hiểm.

(4) Khiêm tốn. Trong dịch tễ học người ta có câu hiểu nôm nà là “không có mô hình toán học nào có thể xác định được diễn biến của dịch, chỉ có con virus quyết định diễn biến của dịch.” Điều này có nghĩa là mô hình phải có yếu tố ‘Vô Thường’, hiểu theo nghĩa uyển chuyển thích ứng với tình huống mỗi ngày. Quên câu nói này có thể dẫn đến nhiều hậu quả khó lường.

Trong khi chờ mô hình Việt Nam, các bạn có thể dùng mô hình của Úc để có một cái nhìn ngắn hạn về diễn biến dịch ở Việt Nam tại đây:

<https://covid19forecast.science.unimelb.edu.au>

[1] Một trong những chức danh của tôi là Professor of Predictive Medicine. Dịch nôm na sang tiếng Việt thì Predictive Medicine là Y khoa Tiên lượng. Nhưng trong thực tế thì không hẳn như thế, danh xưng đó có nghĩa là Y học Cá nhân hoá (Personalized Medicine hay gần đây là Precision Medicine). Y học Cá nhân hoá có nghĩa là bộ môn khoa học có mục tiêu khám phá gen và sử dụng gen để tiên lượng nguy cơ bệnh tật và dự báo hiệu quả của điều trị cho mỗi cá nhân. Đó chính là lí do tại sao tôi dành nhiều thì giờ nghiên cứu về tiên lượng y khoa.

[2] <https://github.com/mrc-ide/covid-sim>

[3] <https://www.nature.com/articles/d41586-020-01685-y>

[4] Mô hình của IHME: <http://www.healthdata.org/covid>

[5] <https://covid19forecasthub.org> hay trang này cũng có nhiều mô hình để tham khảo:

<https://www.aha.org/guidesreports/2020-04-09-compendium-models-predict-spread-covid-19>

[6] <https://www.jmir.org/2021/6/e24285>

[7] <https://doi.org/10.25561/77482>

[8] https://www.doherty.edu.au/uploads/content_doc/McVernon_Modelling_COVID-19_2.pdf

[9]

<https://tiasang.com.vn/khoa-hoc-cong-nghe/george-box-va-hanh-trinh-den-khoa-hoc-thong-ke-3496>

[10]

<https://nguyenvantuan.info/2020/11/13/gioi-thieu-sach-moi-mo-hinh-hoi-qui-va-kham-pha-khoa-hoc>

[11] <https://nguyenvantuan.info/2021/08/07/dich-vu-han-va-van-de-du-lieu>

<https://nguyenvantuan.info/2021/08/24/tat-ca-mo-hinh-tien-luong-covid-deu-sai/>