

真実がわからない中で過去からの学びをどう活かすか

坂巻 顕太郎 (横浜市立大学)

本稿を書いている 2021 年 1 月末は、COVID-19 感染拡大防止のために発出された緊急事態宣言の真っ只中です。期待のワクチンは、不手際から日本では接種が遅れており、少し先の未来すら不透明な状況です。日本では、おそらく 2020 年 2 月のダイヤモンドプリンセス号での集団感染が、COVID-19 に関する議論をはじめの契機になったと思われます。以降、K 値、アビガン、ポピドンヨード、GoTo など、統計的な意味での予測や因果に関する議論を蔑ろにしたものが社会的には話題になりました。その一方で、記述疫学に基づく議論も蔑ろにされ、予測や因果の観点がないと批判する界限も散見されました。Evidence-Based Medicine や Evidence-Based Policy Making が空虚に感じられる社会において、計量生物学の未来はどうなるのでしょうか。機械学習や AI など含むデータサイエンスの領域、計量経済学や計量社会学といったその他のドメインで発展した統計学の領域、疫学も含めた様々な領域とクロスオーバーすることが不可避である中で、計量生物学の役割とは何でしょうか。

私自身は計量生物学の未来に対する答えを持っていないのですが、未来に向けてできることは、過去から学び、アップデートすることしかないと思います。例えば、生存時間解析と言えば、Kaplan-Meier 法、log-rank 検定、Cox 回帰の三つを思い浮かべる人が多かったと思います。今では、比例ハザード性の仮定から逸脱する事例が多くなり、重み付き Log-rank 検定や RMST (restricted mean survival time) などを思い浮かべる人もいるでしょう。一方で、ハザード比についてすべて理解したかと言われると私は自信がありません。2020 年度統計関連学会連合大会では、「評価項目の大小関係に着目して治療効果を評価する統計手法」に関する発表をしました。その中で、Mann-Whitney タイプの効果指標とハザード比との関連に言及し、エンドポイントの分布をどう評価するかという話をしました。個人効果の要約に関連して Hand のパラドックスは説明しましたが、点治療と時間依存性治療の話などはしておらず、ハザードの計算からハザード比を解釈する人には誤解を与えたかもしれません。話が逸れましたが、ここで述べたいことは、アップデートは一つではないということです。RMST という別の方法を用いることもあれば、net benefit や generalized pairwise comparison などの Mann-Whitney タイプの効果指標を用いたハザード比の発展を考えることもできます。過去から学ぶことはそれ自体がとても複雑で、それに基づくアップデートが未来にどう繋がるかはわからないわけですが、選択と集中はせず、様々な観点を持つことの重要性を生存時間解析の例は示していると思います。

過去からの学びに関連して、計量生物学での過去の仕事を少し宣伝させてください。昨年、「ベイズ流決定理論を用いる臨床試験：効用とサンプルサイズ設計」(坂巻他, 2020) という総説が出ました。中では、検定 (p 値) の問題やベイズ流決定理論の有用性を論じています。p 値については、佐藤俊哉先生が中心となって翻訳された ASA 声明 (Wasserstein et al., 2016) にある通り、誤解や誤用が問題視されています。一方で、p 値に関する議論は未だ止むことがありません。その理由の一つは、状況に応じた議論が必要だからだと思います。実際、私自身は多重比較に関する仕事 (坂巻, 2015; 吉田他, 2015) もしており、いまのところ、p 値が絶対悪だとは思っていません。しかし、すべての状況で常に p 値が有用なはずはなく、ベイズ流決定理論を用いる意思決定が重要となる場面が数多くあると考えています。意思決定が

必要な場面における統計学の役割に関する議論は、ニュースレター第 134 号の大橋靖雄先生の巻頭言でのコメントが記憶に新しいですが、私もこの点を考えることが重要だと思います。例えば、近年、リアルワールドエビデンスの議論が盛んになっており、ヒストリカルデータの活用が議論されています。「臨床試験におけるヒストリカルコントロールデータの利用」(武田他, 2015) という総説を出していますが、このような方法を用いる際に p 値を用いる必要はないと思います。p 値については、一方で、Benjamin 先生や Berger 先生のように有意水準を 0.005 に変更するアップデートを考えることもできます。様々な意見がある中で、未来がどうなるかはわかりませんが、「イベント予測モデルの評価指標」(篠崎他, 2020) も含めて、我々の総説が未来の土台の一つになればと思います。

上述のように、過去から学び、様々なアップデートが必要と考えるのは、真実が分からないからです。真実が分かっているならば、最適なものを選択することができます。しかし、現実是不確実なことしかわかりません。そういう状況で生きる統計学はとても面白い学問だと思います。強化学習でも用いて最適なものを選択できればいいのですが、報酬 (reward) が定まらない私のような人間には最適を決めることは難しいです。多重比較は予測や推定に関する発展があり、複数のエンドポイントについては複合エンドポイントや競合リスクに関する発展があります。ベイズ流決定理論は抗癌剤開発における意思決定に少しずつ応用されている印象があります。今までの学びを活かして、自分なりに右往左往しながら研究を進めることで、計量生物学の未来に寄与できればと思います。

参考文献

- 坂巻頭太郎. (2015). 構造化仮説群の多重性制御法 1 (Gatekeeping 法). 計量生物学 **36**, S147-S164.
- 坂巻頭太郎, 兼清道雄, 大和田章一, 松浦健太郎, 柿爪智行, 高橋文博, 高沢翔, 萩原駿祐, 森田智視. (2020). ベイズ流決定理論を用いる臨床試験: 効用とサンプルサイズ設計. 計量生物学 **41**, 55-91.
- 篠崎智大, 横田勲, 大庭幸治, 上妻佳代子, 坂巻頭太郎. (2020). イベント予測モデルの評価指標. 計量生物学 **41**, 1-35.
- 武田健太郎, 大庭真梨, 柿爪智行, 坂巻頭太郎, 田栗正隆, 森田智視. (2015). 臨床試験におけるヒストリカルコントロールデータの利用. 計量生物学 **36**, 25-50.
- Wasserstein, R.L., Lazar, N.A. (2016). Editorial: The ASA's statement on p-values: Context, process, and purpose. *The American Statistician* **70**, 129-133. 佐藤俊哉 訳. 統計的有意性と P 値に関する ASA 声明. 日本計量生物学会. <http://www.biometrics.gr.jp/news/all/ASA.pdf>. (Accessed February 1, 2021).
- 吉田征太郎, 坂巻頭太郎. (2015). 臨床試験における Gatekeeping 法の適用例. 計量生物学 **36**, S165-S171.