

## 予測モデル比較のための評価指標

東京大学大学院医学系研究科  
健康科学・看護学専攻生物統計学分野  
篠崎 智大

**背景：**診療や健診・検診におけるリスク予測モデルの有用性は言を俟たないが、役立ちそうな生体指標が新しく注目されてきたり、既存の予測モデルが目の前の集団と生物学的・社会的特徴の大きく異なる集団から開発されたりした場合には、新規にモデルを構築・評価する努力が必要となる。Akaike (1974) の統計的枠組に見られるように、予測モデルはその構造が誤特定されていても十分に有用たりうるが、実際上は単一の基準 (eg. AIC) のみでモデル比較を行うわけにはいかない; 現代の医学雑誌では少なくとも、イベント判別 (discrimination)、リスク較正 (calibration)、異なるモデルによるリスク再分類 (reclassification) の指標を報告することが求められる。最近は教室で予測モデルの話題が出るのが少ないので、いまや常識のように求められるこれらの指標を説明する。

**方法：**疫学・臨床研究分野での予測モデル評価で著名な Pepe, Cook, Steyerberg, Pencina らの論文を簡単にまとめた。今回紹介する指標はいずれも簡単な修正を施せば生存時間データに適用できる。発表者が行った解析事例を、SAS プログラム例を併せて紹介する。

**結果：**データ解析の常ではあるが、予測モデル評価の文脈においても、どれか単一の指標が他の指標に秀でていたということはなく、現状では複数の評価指標を報告して読者に異なる側面を見てもらうのがよいであろう。データ解析の例では、ある予測指標ではモデル予測力の改善がほぼ見込まれないが、他の指標では旧モデルを凌駕し、臨床的にも提案に値する結果と考えられた。

**考察：**臨床家や疫学者に理解しやすい (と思われている) 指標はいろいろ提案されて使われてはいるものの、我々が「計算できればいいや」では困るわけで、十分に検討されていない統計的・臨床的性質の批判的検討の理解が大切である (eg. Hilden, 2014)。今後は、より複雑な状況 (たとえば予測変数の経時測定データや競合リスクが問題になる場合など) への拡張はもちろん、今回取り上げなかった (必ずしも金銭的なものだけではない) コストを考慮した意思決定理論へ展開していくことは容易に想像が付き (Greenland, 2008; Baker et al., 2014)、テーマとしても面白いかもしれない。

## 文献

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **19**, 716–723.

Baker, S.G., Schuit, E., Steyerberg, E.W., Pencina, M.J., Vickers, A., Moons, K.G., Mol, B.W. and Lindeman, K.S. (2014). How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through. *Statistics in Medicine*, publish online.

Greenland, S. (2008). The need for reorientation toward cost-effective prediction (comment). *Statistics in Medicine*, **27**, 199–206.

Hilden, J. (2014). Commentary: On NRI, IDI, and "good-looking" statistics with nothing underneath. *Epidemiology*, **25**, 265–267.