

予測モデル（ロジスティック回帰）におけるパラメータ推定と評価指標の補正について

リスク因子・症状等から現在の病状を予測（diagnostic model）、あるいは未来のイベントを予測（prognostic model）する予測モデルに関する研究を報告するガイドラインとして、2015年に Transparent Reporting of multivariable prediction model for individual Prognosis Or Diagnosis (TRIPOD)が発表された¹⁾。これは、この解説の中で Collin らが述べているように、近年予測モデルを構築、あるいはモデルの validation や再構築、それら両方を目的とする研究が増加しているにもかかわらず、その報告の質が低いものが多いことによる。

TRIPOD のチェックリストには、予測因子（predictor）の選定を含む予測モデルの構築方法と、モデルの予測能力を測る指標とその値を提示するよう求められる項目がある。解説では、予測モデルを構成する predictor の事前の候補とした理由・選定した方法の他、実際のモデル構成時にどのような方法で選択したか（forward や backward selection などの自動選択方法など）を報告するよう求めている。また、モデルの予測能としては、モデルによる予測と実際に観測されたアウトカムとの一致度を測る較正 (Calibration)、イベントの有無の判別能力 (discrimination) を報告すべきとしている。Calibration として、予測されたリスクの確率を X 軸に、実際の観測されたアウトカムを Y 軸にグラフ化する Calibration plot を例としているが、他にも Hosmer-Lemeshow 検定や Calibration-in-the-large、Calibration slope などがある。Discrimination として最も一般的に使用される指標は C 統計量であり、これは ROC 曲線下面積と一致する。しかし、予測モデルのレビューにおいて、predictor の選択方法や calibration、Discrimination の報告のあるものは多くはなく、predictor の選択方法では 36%、Calibration では 26%、Discrimination では 44% としている研究がある。尚、これら以外にも全体のパフォーマンスを測る指標としての R^2 や Brier score、Reclassification として Reclassification table、Reclassification statistics、Net reclassification index (NRI) などもある。

こうしたモデルにおける predictor の選択・予測能力を測る際に留意する必要があるのが overfitting の問題であり、これを考慮・補正する方法として、モデルの構成時に予測モデルの回帰係数をより小さく推定する shrinkage 法、回帰係数や予測能力の指標を internal validation により補正する方法²⁾などがある。本抄読会ではこれらについてまとめ、いくつかの手法で第二世代薬物溶出性ステント留置後の実データをあてはめた結果を報告、それぞれの特徴について考察する。

主要参考文献

1. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj*. 2015;350:g7594.
2. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-81.