Artificial Neural Networks Outperform Linear Regression in Estimating Patient-Reported Outcomes Following Upper Extremity Fractures with a Larger Number of Included Variables

Niels Brinkman, BS*; Romil Shah, MD; Stephen Gwilym, DPhil; David Ring, MD, PhD; Prakash Jayakumar, MD, PhD; Job N. Doornberg, MD, PhD*

Purpose: Symptom intensity and level of capability following upper extremity fracture (UEF) can be estimated by analyzing mental, social, and physical health at baseline or initial weeks after injury. Such estimates can help surgeons and patients with shared decision-making and expectation management during recovery. We compared the performance of linear regression (LR) and artificial neural network (ANN) models in estimating 9-month patient-reported outcomes (PROs) after UEF. We also evaluated differences in performance based on the inclusion of different numbers of explanatory variables.

Methods: We studied 731 patients with isolated shoulder, elbow, or wrist fracture who completed demographics, measures of mental and social health, and PROs at baseline, 4-6 weeks, and 6-9 months post-injury. PROs included PROMIS (Patient-Reported Outcomes Measurement Information System) Upper Extremity (UE) Physical Function (PF), PROMIS Pain Interference (PI), QuickDASH (an abbreviated version of the Disabilities of the Arm, Shoulder and Hand [DASH] questionnaire), and Likert Pain Scale (LPS). We developed ANN and LR models with different numbers of explanatory variables (20, 23, 29, 34, 54) to predict 9-month PROs. The predictive performance was measured as the accuracy of the predicted value being within 1 minimal clinically important difference of the actual 9-month PRO value. The 29-variable model was chosen as primary outcome to optimize model performance, limit the risk of model overfit, and maintain clinical relevance of the included variables.

Results: The accuracy of ANN versus LR models in the primary model was 83% versus 73% in QuickDASH, 68% versus 65% in PROMIS UE PF, 66% versus 62% in PROMIS PI, and 78% versus 65% in LPS. The performance of ANN and LR models generally improved with a greater number of explanatory variables, in which the ANN models seemed to benefit most. LR performed better in all 20-variable models and the 23-variable model for LPS. Notably, mental and social health factors demonstrated relatively high variable importance in both ANN and LR models.

Conclusion: ANNs can outperform LRs in estimating 9-month PROs after UEF, especially when a larger number of variables were included. LR seems to perform better in models with a smaller number of included variables. Further research is needed to define the optimal number and type of variables for ANNs to ensure variable selection is both feasible and meaningful for informing clinical practice.