



Commentary



Commentary on “Development and Validation of an Online Calculator to Predict Proximal Junctional Kyphosis After Adult Spinal Deformity Surgery Using Machine Learning”

In Ho Han

Department of Neurosurgery, Pusan National University Hospital, Pusan National University School of Medicine, Busan, Korea

Corresponding Author

In Ho Han

<https://orcid.org/0000-0001-7193-6533>

Department of Neurosurgery, School of Medicine, Pusan National University Hospital, 179 Gudeok-ro, Seo-gu, Busan 49241, Korea

Email: farlateral@hanmail.net

See the article “Development and Validation of an Online Calculator to Predict Proximal Junctional Kyphosis After Adult Spinal Deformity Surgery Using Machine Learning” via <https://doi.org/10.14245/ns.2342434.217>.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2023 by the Korean Spinal Neurosurgery Society

Recently, research on artificial intelligence (AI), the most important element of the Fourth Industrial Revolution, has been actively conducted in the medical field. Medical AI is mainly applied to predict treatment outcomes or medical image diagnosis.^{1,2} In the field of spinal surgery, in particular, there is a lot of interest in developing a machine learning (ML) model that predicts the outcome of patients after spinal surgery.³ Regarding adult spinal deformity (ASD), studies were conducted to predict postoperative complications, mortality, and length of hospital stay.^{4,5} In addition, a ML model that automatically measures spinopelvic parameters in spinal radiographs was also introduced.³

Many AI studies that predict postoperative outcomes adopt a retrospective approach.⁶ They retrospectively organize data from patients who underwent previous surgeries, using it as training data. Subsequently, they compare the performance of existing algorithms or modified algorithms in predicting patient outcomes such as surgical results, complications, and mortality.

This study⁷ followed a similar research process. Instead of inputting as many variables as possible, as seen in other studies, this study found the independent prognostic factor through a statistical approach to reduce the number of variables in the input data.

The study⁷ went a step further by creating an online calculator program based on the predictive model. This program is available to other surgeons for easy and free access. By entering eight data points (type of deformity, age at the time of surgery, body mass index at the time of surgery, Scoliosis Research Society (SRS)-curve pattern, SRS-PI-LL (pelvic incidence minus lumbar lordosis) modifier, SRS-global balance modifier, pelvic incidence at baseline, and proximal junctional angle at the postoperative state) surgeons can promptly confirm the probability of proximal junctional kyphosis (PJK) after ASD surgery. The minimal input items and the ease of data entry through a click interface were designed to enhance user convenience. This feature distinguishes it from other ML studies.

In contrast to the typical approach of AI models yielding results in a yes or no or A, B, C, D classification, the online calculator adopts a method of deriving probability values. Instead of making a binary decision like ‘PJK will occur’ or ‘will not occur,’ it produces results stat-

ing, for example, that the patient’s PJK risk is 13.81%. This approach shares similarities with tools like FRAX, used for calculating the risk of osteoporotic fractures.⁸ When provided with key risk factors for osteoporotic fractures, FRAX calculates and presents the individual’s 10-year fracture probability through the interactions among these risk factors.

Recently, there has been a growing trend in AI research focused on predicting the treatment outcomes for patients like this. It is crucial to engage in fundamental discussions regarding these studies. Firstly, is the training data used for the predictive model both qualitatively and quantitatively satisfactory? The data used for training AI models should be of high quality and quantity, aligning with the intended purpose, and ideally reflecting clinical realities because ultimately, AI operates based on training data.⁶ Therefore, the process of constructing training data is considered one of the most crucial stages in ML research. It would be helpful to provide an example using FRAX, the fracture risk assessment tool mentioned earlier. FRAX was developed based on large-scale cohort study results in Europe, North America, Asia, and Australia.⁸ It evaluated the absolute risk of fractures by analyzing data from a total of 60,000 participants, including 5,400 fracture cases, and 1,000 hip fractures, from 12 prospective cohort studies.⁸ Although FRAX has some limitations, it is presently employed as a crucial tool for predicting fractures in clinical practice for osteoporosis treatment. In this context, there are limitations in both quantitative and qualitative aspects of the data in the study. Out of 201 ASD patients undergoing surgery over 10 years across 16 centers, only 49 of them experiencing PJK. The training data for ML appears limited considering recent trends in AI research. Additionally, surgical outcomes may be influenced by the medical center or surgeon, with the possibility that more than 16 surgeons could impact the occurrence of PJK. However, if the research group consistently and accurately maintains a database for ASD surgery patients and continuously updates the calculator, it could become an excellent program.

Secondly, in the realm of recent advancements in medical AI, a pivotal focus lies on the validation of AI systems.⁶ Typically, the validation of AI performance is divided into internal and external validation. Internal validation involves a simple assessment of the performance of AI algorithms during their development, often leading to a tendency to overestimate performance. In this study, a commonly employed method of internal validation, namely cross-validation, was used. However, recent AI research emphasizes the importance of external validation. External validation refers to evaluating the performance of AI using

data collected independently of the initial dataset. Unfortunately, in this study, external validation was not employed. The use of the term “validation” in the paper’s title could potentially mislead readers into thinking that the AI program underwent external validation. It is speculated that external validation was challenging due to the already collected patient data from multiple institutions and the relatively low frequency of ASD surgeries. Furthermore, the occurrence of PJK was evaluated at the final follow-up of each patient. The final follow-up time varied among patients. The assessment timing for determining PJK occurrence is vague, and the criteria for external validation also seem ambiguous. Generally, PJK is known to occur relatively shortly after surgery, often within the first few months postoperatively.⁹ Therefore, it seems necessary to establish an expected occurrence time, similar to examining the 10-year probability of fracture in FRAX. Considering these uncertainties, careful consideration is needed for future external validation of this AI system.

The third issue is whether the AI program is genuinely beneficial in practically assisting patient treatment. While an AI program may demonstrate outstanding performance, it may essentially be of little use if it does not contribute significantly to patient diagnosis or treatment. To assess the practical utility of AI, randomized clinical trials comparing a group using the AI program with a control group are necessary. However, examples of AI validated for such clinical utility are rare to date.⁶ In this study, the authors concluded that predicting PJK probability through the online calculator would assist in formulating subsequent treatment strategies. So, if PJK probability is confirmed using this program, how can it be beneficial for the patient’s prognosis? For instance, if the PJK probabilities immediately after surgery are 75%, 50%, and 25%, how should treatment strategies be approached for each individual patient? The results of AI predicting a patient’s prognosis should not be dismissed as a mere ‘So what?’. Medical AI predicting surgical outcomes requires additional research to demonstrate future clinical utility.

In summary, the study developed a predictive model for PJK but distinguished itself by creating an online calculator for PJK probability for surgeons. Despite data limitations, it shows promise for medical AI research. Future improvements are expected for enhanced findings and a better AI program benefiting patients and surgeons. In future spine surgery AI research, we hope for an advanced program considering both qualitative and quantitative data, alongside external validation and clinical utility.

• **Conflict of Interest:** The author has nothing to disclose.

REFERENCES

1. Noh SH, Cho PG, Kim KN, et al. Artificial intelligence for neurosurgery: current state and future directions. *J Korean Neurosurg Soc* 2023;66:113-20.
2. Kim M, Yun J, Cho Y, et al. Deep learning in medical imaging. *Neurospine* 2019;16:657-68.
3. Joshi RS, Haddad AF, Lau D, et al. Artificial intelligence for adult spinal deformity. *Neurospine* 2019;16:686-94.
4. Arora A, Demb J, Cummins DD, et al. Predictive models to assess risk of extended length of stay in adults with spinal deformity and lumbar degenerative pathology: development and internal validation. *Spine J* 2023;23:457-66.
5. Noh SH, Lee HS, Park GE, et al. Predicting mechanical complications after adult spinal deformity operation using a machine learning based on modified global alignment and proportion scoring with body mass index and bone mineral density. *Neurospine* 2023;20:265-74.
6. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
7. Lee CH, Jo DJ, Oh JK, et al. Development and validation of an online calculator to predict proximal junctional kyphosis after adult spinal deformity surgery using machine learning. *Neurospine* 2023;20:1272-80.
8. Kanis JA, McCloskey EV, Johansson H, et al. Development and use of FRAX in osteoporosis. *Osteoporos Int* 2010;21 Suppl 2:S407-13.
9. Lee JW, Park YS. Proximal junctional kyphosis: diagnosis, pathogenesis, and treatment. *Asian Spine J* 2016;10:593-600.