Neurospine

Commentary

Predicting Proximal Junctional Kyphosis After Adult Spinal Deformity Surgery: A Step Towards True "Precision" Medicine?: Commentary on "Development and Validation of an Online Calculator to Predict Proximal Junctional Kyphosis After Adult Spinal Deformity Surgery Using Machine Learning"

Lara M. Höbner, Alexandra Grob, Victor E. Staartjes

Machine Intelligence in Clinical Neuroscience & Microsurgical Neuroanatomy (MICN) Laboratory, Department of Neurosurgery, Clinical Neuroscience Center, University Hospital Zürich, University of Zürich, Zürich, Switzerland

Corresponding Author

Victor E. Staartjes
https://orcid.org/0000-0003-1039-2098

Group Leader MICN Laboratory, Department of Neurosurgery, University Hospital Zürich, Frauenklinikstrasse 10, Zürich 8091, Switzerland
Email: victoregon.staartjes@usz.ch

In the past decades, a deeper understanding of the biomechanical underpinnings of adult spinal deformity (ASD) surgery has been gained, with an increased focus on global sagittal alignment. Still, the development of proximal junctional kyphosis (PJK) or even proximal junctional failure after correction remain a relatively frequent and clinically relevant occurrence, and represent major drivers of reoperations and morbidity.[1]

Lee et al.[2] approach the problem of PJK by aiming at reliably predicting its risk. With this, they aim to promote enhanced patient counseling and risk-benefit management, but also allow for refined therapeutic approaches – corrections could in theory be "personalized" to every single patient, even more so than they are anyways already, by considering personalized risk profiles. Moreover, specifically for PJK, postoperative patient-specific evaluations of the correction could identify those who may benefit from early revision surgery. The authors apply data from a large multi-institutional database from 16 Korean centers, strengthening the potential generalizability of their approach. In total, 201 patients with a minimum follow-up of 1 year were included, of which 49 (24.4%) experienced PJK – which was defined as a proximal junctional angle (PJA) of 20° or greater, or as an increase in PJA of 10° or greater. All patients were then randomly split into train and test sets, hyperparameters were tuned via a cross-validation approach, and a range of machine learning (ML) techniques applied. Input parameters of the final model, which was based on the random forest algorithm, include age and body mass index (BMI), deformity etiology (idiopathic, degenerative, neuromuscular, etc.), curve type, and pelvic as well as global parameters. In addi-

tion, directly postoperative PJA is included as an input: This inclusion of a postoperative parameter makes the current model less suitable as a preoperative predictive tool, although this parameter is of course one of the most important independent risk factors. This is corroborated by the separate multivariable analysis that the authors have performed to identify independent risk factors for PJK (directly postoperative PJA, BMI, etiology), in which age and bone mineral density intererestingly were not independently predictive of PJK. The final model has been incorporated into a freely accessible web-app, and was able to predict PJK with an area-under-the-curve of 0.76 at internal validation.

Clinical prediction modeling has swiftly become one of the most frequent applications of ML in medicine: The ability to predict the future would certainly benefit patients and surgeons. Although the performance of ML-based predictions is often assumed to be superior to more traditional statistical approaches such as generalized linear models (logistic or linear regression), this is in fact overall not the case for tabulated medical data.[3] Tabulated medical data usually do not show a very high dimensionality, nor a patient number in the millions—for these reasons, it is preferable to apply less complex ML architectures that are also natively interpretable for most medical applications—deep learning, for example, should not play a role for tabulated medical data. This has been implemented rather well by Lee et al.,[2] who use a very reasonable set of architectures considering the underlying data. In a future step, it might be interesting and useful to not only include premeasured, tabulated data, but to supplement radiographs, computed tomography, or magnetic resonance imaging directly: On the one hand, automated extraction of the parameters that are necessary for prediction may be more efficient and has already been shown to be feasible.[4] On the other hand, radiomic feature extraction could reveal additional predictors of PJK that are not currently or routinely captured.

Because clinical decisions are bound to be made from published and accessible prediction models—as is the case here—it is vital to maintain high methodological standards in their development and ensure rigorous external validation before deploying models.[5] The authors apply a high standard of ML methods. One strength of this particular paper certainly is the wide range of participating centers (albeit—as the authors acknowledge—currently only confined to Korean centers) and the high quality of data collection. The high number of centers and participating surgeons goes a long way towards ensuring generalizability. The current validation strategy with random splitting and internal validation shows promising results, but only after true external validation should clinical prediction models be adopted into daily clinical practice. It may also be interesting to look at calibration of the predicted risk for PJK (how well do the predicted risks correspond to the truly observed risk?)—instead of only at discrimination (how good is the binary prediction?), as calibration is often viewed as more critical to clinicians than discrimination performance.

Still, many prediction tasks appear simply too difficult in the real-world (meaning, with proper external validation)—probably, it is simply unrealistic to expect to be able to predict the future with any great amount of accuracy.[6] Conceptually, this is even clearer in medical prediction modeling: Predicting future outcomes that depend on hundreds of factors would—of course—require collection and integration of these hundreds of factors. Apart from the impracticality of collecting and inputing such wealths of data into an online calculator in daily practice, in medicine, we usually do not even have the case numbers to allow for training of models with hundreds of factors (which would require tens of thousands of patients for proper clinical prediction modeling). In turn, such amounts of training data would also massively increase the risk of overfitting. For all of these reasons, clinical prediction modeling is usually confined to only a dozen or so inputs, which of course then limits how accurately predictions of complex medical outcomes can be made.

Clinical prediction modeling is certainly still booming, but even the most well-validated and best-performing prediction models beg one core question: Even if models are shown to work robustly in silico (i.e., they perform well at external validation), do they lead to a tangible or measurable clinical benefit? To the best of the authors' knowledge, no study has as of yet demonstrated any measurable clinical benefit of adding clinical prediction models—this is a question that the next decade of clinical research will have to answer.

In conclusion, the authors present a great step towards personalized/precision medicine with a well-developed web calculator, and it seems likely that such approaches will yield particularly useful results in ASD surgery and even more so in clearly outlined and measurable problems such as PJK—as this is already a highly quantitative domain. We commend Lee et al.[2] for their work, and are confident that further development including rigorous external validation as well as increased automatization using direct extraction of parameters will eventually enable a true clinical benefit to real-world patients.

• **Conflict of Interest:** The authors have nothing to disclose.

REFERENCES

1. Ha Y, Maruo K, Racine L, et al. Proximal junctional kyphosis and clinical outcomes in adult spinal deformity surgery with fusion from the thoracic spine to the sacrum: a comparison of proximal and distal upper instrumented vertebrae: clinical article. J Neurosurg Spine 2013;19:360-9.
2. 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.
3. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22.
4. Nakarai H, Cina A, Jutzeler C, et al. Automatic calculation of cervical spine parameters using deep learning: development and validation on an external dataset. Glob Spine J 2023 Oct 9:21925682231205352. doi: 10.1177/21925682231205352. [Epub].
5. Staartjes VE, Kernbach JM. Foundations of machine learning-based clinical prediction modeling: part III-model evaluation and other points of significance. Acta Neurochir Suppl 2022;134:23-31.
6. Staartjes VE, Stumpo V, Ricciardi L, et al. FUSE-ML: development and external validation of a clinical prediction model for mid-term outcomes after lumbar spinal fusion for degenerative disease. Eur Spine J 2022;31:2629-38.