Optimizing care in lumbar radiculopathy and neurogenic claudication: from injection to inference, and from clinician to algorithm

Verheijen, E.J.A.

11

CONCLUSION AND DISCUSSION

The findings of this thesis contribute to the knowledge of conservative management with transforaminal epidural steroid injections in patients with lumbar radiculopathy, as well as underline the potential of innovations with artificial intelligence in the diagnosis and prognosis of spinal stenosis, marking another step toward more patient-tailored treatment strategies.

BALANCING SURGERY AND CONSERVATIVE CARE IN LUMBAR RADICULAR PAIN: THE ROLE OF EPIDURAL STEROID INJECTIONS

Considerations for surgical intervention

The current national guidelines in the Netherlands recommend that patients presenting to their general practitioner with signs of lumbar radiculopathy undergo an initial period of six to eight weeks of conservative management, consisting of oral pain medication and physical therapy if necessary. If symptoms persist beyond this timeframe, the patient can be referred to the neurologist for further evaluation, including MRI examination. The neurologist may then refer the patient to an outpatient pain clinic or a neurosurgeon if a surgically treatable anatomical substrate is identified.

The majority of patients diagnosed with lumbar radiculopathy demonstrates a clinically relevant disc herniation (LDH) on MRI as confirmed by our retrospective cohort analysis (**Chapter 3**). Previous research has demonstrated that surgical intervention is not necessarily the optimal treatment strategy, as lumbar disc herniation is often a self-limiting condition. A randomized trial involving 283 patients with radicular symptoms of 6 to 12 weeks due to LDH found that early discectomy did not yield superior long-term outcomes compared to prolonged conservative management after one year of follow-up [1]. The mere benefit of surgical intervention was a more rapid resolution of symptoms, while overall treatment costs remained comparable between the two strategies [2]. In clinical practice, one in five patients with disc herniation opt for immediate surgery, whereas the majority – over 50% – choose to continue conservative therapy. Similarly, the natural history of and the relative effectiveness of surgical intervention for lumbar radicular pain due to spinal stenosis remain subjects of ongoing debate.

We demonstrated that a large proportion of patients without disc herniation have a degenerative stenosis as the underlying cause of their symptoms (**Chapter 3**). A 2016 Cochrane Review concluded that there is limited evidence supporting surgical treatment over conservative management in this patient

population, a conclusion largely attributed to the methodological heterogeneity and low quality of existing trials [3]. Yet, a consensus group from the World Federation of Neurosurgical Society Spine Committee in 2020 agreed on the recommendation of decompressive surgery, in particular for patients with moderate-to-severe symptoms, citing evidence that surgical intervention resulted in faster and significantly greater symptom relief than conservative treatment, especially in the long term [4]. However, this recommendation was based largely on an as-treated analysis from a single study that combined a randomized and an observational cohort, and was compromised by a high rate of treatment cross-over [5]. The consensus group acknowledged that conservatively managed patients also demonstrated symptom improvement, albeit at a slower rate. In conclusion, the recommendation for a surgical solution for lumbar radiculopathy due to stenosis has weak strength and there is a need for further high-quality research in this area.

The varying effectiveness of surgical intervention to remove the compressive lesion and the discrepancy between clinical symptoms and radiological findings has fuelled the debate on the true cause of radicular pain. At first, it was assumed that physical neural compromise due to disc herniation and stenosis was the sole cause of symptoms. However, multiple studies have demonstrated the lack of correlation between clinical symptoms and nerve root compression as this is a frequent radiological finding in asymptomatic patients [6-13]. Moreover, symptomatic patients may experience improvement in symptoms without intervention that alters the underlying pathology, or may not show any compressive pathology at all whilst suffering from severe radicular pain. Although a correlation was found between contact pressure and neurological impairment [14], it was suggested that something else instigated radicular pain. Inflammation was implicated in the pathophysiology of lumbar radicular pain rather than physical compression after histological evidence was found of inflamed posterior nerve roots during laminectomy [15]. This was supported by findings that chymopapain, an anti-inflammatory substance used for chemonucleolysis of disc herniation, could cause rapid relief of leg pain preceding any change in disc herniation size or degree of nerve root compression [16, 17]. Furthermore, several additional studies demonstrated the presence of high levels of pro-inflammatory enzymes (PLA_2) and cytokines (IL-1α, IL-1β, IL-6, IL-8, TNF-α and PGE2) in patients with radicular pain with different injured spinal tissues releasing different types of inflammatory factors [18-23]. Additionally, some evidence has suggested that immunological processes play a role in the instigation of radicular pain as antibody levels to cell components of the central and peripheral nervous system and markers of glial cell and nerve damage have been shown to be elevated in patients with disc herniation [24, 25]. Hence, it is now an accepted concept that

lumbar radicular pain is caused by a complex interplay of inflammation, immune factors and mechanical compression of the nerve root.

Conservative management: the promise and limitations of epidural steroid injections

Considering that, on average, conservative therapy may be as effective as surgery in the long term for patients with lumbar radicular pain, it represents a viable alternative – provided that the patient's symptoms can be adequately reduced. Patients presenting with lumbar radiculopathy are initially treated with oral analgesics, starting with paracetamol and non-steroidal anti-inflammatory drugs (NSAIDs). However, a substantial proportion of patients are prescribed additional opioid analgesics. Our retrospective data indicate that 57% of patients presenting to the outpatient pain clinic were using some form of morphine-like analgesics (**Chapter 3**). In light of the ongoing opioid crisis and the well-documented risks of dependency and adverse effects associated with chronic opioid use, this reliance on opioid analgesia is undesirable. Moreover, despite this pharmacological treatment, many patients continued to seek further therapeutic options.

Epidural steroid injections (ESI), although not addressing the underlying pathology directly – they do not create space for the neural elements or remove the herniated disc – are intended to modulate the inflammatory and immunological responses assumed to contribute to radicular pain. Corticosteroids are primarily thought to inhibit PLA_2, an important enzyme in the pro-inflammatory cascade, resulting in the reduced excretion of hyperalgesic prostaglandins, thromboxanes, and leukotrienes, thereby alleviating symptoms and creating a more tolerable clinical condition for the patient [23, 26, 27]. Hence, when effective, ESIs constitute a valuable addition to the spectrum of non-surgical treatment modalities. Conversely, in cases where ESIs fail to provide sufficient symptom relief, patients may still proceed to surgical intervention.

The use of epidural steroid injections (ESIs) has seen increasing prevalence, particularly as the role of inflammation in the onset of radicular pain has become more widely recognized. The current literature, however, is not unequivocal regarding the effectiveness of ESI in patients with lumbar radiculopathy. Interestingly, despite the lack of evidence, this treatment has become increasingly popular over recent decades. Our meta-analysis comparing the effectiveness of ESI to placebo injections provides some evidence that ESI is superior as this group reported greater reductions in pain scores and improvement of physical functioning (**Chapter 2**). Though, this was limited to short-term follow-up and the minimally clinical important difference (MCID) was not met in every study.

The interpretation of these findings is further complicated by substantial heterogeneity across studies, as ESIs can be administered via different approaches (caudal, interlaminar, or transforaminal) and involve varying corticosteroid formulations and dosages.

Optimizing patient selection: who benefits most from transforaminal epidural injections?

The paucity of definitive evidence supporting the effectiveness of transforaminal epidural steroid injections (TEI) may, in part, stem from suboptimal patient selection. Hence, identifying prognostic factors to guide clinicians in determining which patients are most suitable candidates for TEI is essential. However, this objective is impeded by the dearth of comparable studies. In our meta-analysis, sensitivity analyses were performed to explore potential determinants of treatment response, suggesting that ESI may be more effective in patients with MRI-confirmed LDH compared to those diagnosed solely on clinical assessment (**Chapter 2**). However, due to the necessity of pooling all three epidural approaches and the limited number of studies included in each analysis, these findings should be interpreted with caution.

Contrastingly, our large retrospective cohort study demonstrated that the short-term effectiveness of TEI is independent of the presence of a disc herniation on MRI (**Chapter 3**). This is an interesting finding, as it challenges the prevailing but unsubstantiated assumption that TEI is primarily beneficial for patients with LDH. While LDH generally follows a self-limiting course with eventual spontaneous resolution, spinal stenosis caused by degenerative changes typically progresses over time. Given that the effects of corticosteroids after a single injection are expected to diminish after several weeks, one could argue that TEI may serve as a temporary bridge to spontaneous recovery in LDH patients, whereas in those with spinal stenosis, symptom relief may be transient, but symptoms will eventually recur. However, as mentioned before, spontaneous symptom resolution can also occur in patients with spinal stenosis, and the precise role of inflammation – as well as the extent to which TEI modulates this process – is uncertain. Interestingly, in our systematic review of prognostic factors for TEI, two studies suggested that patients with spinal stenosis, specifically those with foraminal stenosis, may benefit more than LDH patients (**Chapter 4**). A potential explanation is that degenerative stenosis causes more dynamic, movement-dependent compression and recurrent cycles of inflammation around the nerve root. These intermittent flare-ups of perineural inflammation and vascular congestion may be more amenable to corticosteroid treatment than the continuous inflammation and static compression from disc herniation. In addition, in foraminal stenosis, corticosteroids delivered to the

neuroforamen may be confined by the stenotic components, hence, increasing the local drug concentration around the affected nerve. In contrast, in central or lateral recess stenosis, the injectables have to diffuse reducing their localized anti-inflammatory effects. Furthermore, vascularization is higher centrally and in the lateral recess, where corticosteroids may be more rapidly absorbed into the circulation, whereas in foraminal stenosis, compression-induced venous congestion and impaired perineural circulation may promote corticosteroid accumulation, prolonging its anti-inflammatory effects. Finally, a shorter symptom duration was associated with improved outcomes, although this relationship was identified in only four out of ten studies, and, hence, cannot be considered an established prognostic variable. The inconsistency of these findings highlights the need for studies specifically designed to establish reliable predictors of TEI effectiveness.

Biomarkers of inflammation

Since administration of corticosteroids is aimed at attenuating neuroinflammation, it is reasonable to postulate that TEI would be most beneficial for patients in whom inflammatory mechanisms play a predominant role. Consequently, identifying reliable biomarkers indicative of increased inflammatory activity is imperative for optimizing patient selection for TEI.

Gadolinium-based contrast agents are frequently employed in MRI examinations to enhance visualization of neuroinflammatory processes. Theoretically, the presence of gadolinium enhancement in the nerve root could serve as a predictor of TEI effectiveness. This hypothesis is supported by a retrospective study reporting greater symptomatic improvement following TEI in patients with nerve root enhancement on MRI compared to those without contrast enhancement [28]. However, histological analyses of intervertebral disc samples obtained from surgical patients have not demonstrated significant differences in macrophage-mediated inflammation between contrast-enhanced and non-enhanced groups [29].

Another proposed imaging biomarker is the presence of Modic changes (MC) on MRI, which are postulated to represent vascular deficiency, inducing an inflammatory milieu in the vertebral endplates. MC can be classified into three types, although only two are commonly observed. Type 1 MC is thought to represent an acute inflammatory process characterized by bone marrow oedema, whereas Type 2 MC is associated with a more chronic inflammatory state involving fatty marrow changes [30-32]. Yet, there is evidence suggesting that type 1 can develop into type 2 over time. Studies have correlated the presence of MC with an inflammatory state of the intervertebral disc and altered macrophage

distribution, with a relative depletion of anti-inflammatory M2 macrophages compared to pro-inflammatory M1 macrophages [33]. Since corticosteroids have been demonstrated to drive macrophage differentiation towards the M2 type in a mice model with lung injury [34], TEI could hypothetically facilitate an anti-inflammatory shift. However, our prospective cohort study found no significant differences in patient outcomes following TEI between those with and without MC (**Chapter 7**). Nevertheless, these results must be interpreted with caution as there was a preponderance of MC type 2 (93.5%). Moreover, it is possible that in patients with MC type 2, recurrent inflammatory episodes occur – resulting in oedema (MC type 1) – but are not detectable as type 1, since fatty marrow transformation has already taken place, leading in a MC type 2 score. This suggests that MC classification may not reliably distinguish between active and past inflammation. We recommend that this correlation should be repeated in a larger study with a more balanced representation of MC types. Among MRI features, mean spinal nerve intensity from axial Dixon T2-weighted water-only images has been correlated with patient outcome after TEI, with higher nerve intensity being associated with greater symptomatic relief [35]. The authors suggested that axial T2-weighted fat-saturated MRI scans can detect perineural inflammatory and hyperaemic changes that may not be visible on routine T1- and T2-weighted imaging used for MC classification.

Beyond conventional MRI, positron emission tomography combined with magnetic resonance imaging (PET/MRI) has emerged as a promising modality for detecting inflammatory changes. This imaging technique makes use of radioligands targeting inflammatory markers to visualize neuroinflammation. A study by Albrecht et al. demonstrated elevated levels of the neuroinflammation marker 18-kDa translocator protein (TSPO) in the neuroforamen of patients with chronic radicular pain compared to healthy controls, with a spatial distribution corresponding to the side of leg pain [36]. Moreover, increased TSPO expression correlated with improved outcomes following ESI, although the number of patients receiving an injection in this study was limited. Similarly, Lutke Schipholt et al. used this imaging modality to quantify the degree of inflammation in the spinal cord and neuroforamen in patients that suffered from cervical radiculopathy, using unaffected neuroforamina within the same patient as control, and found elevated inflammation levels at the affected neuroforamina [37]. In a subsequent study, they confirmed these findings by demonstrating higher levels of inflammation in neuroforamina corresponding to symptomatic cervical radiculopathy when compared to asymptomatic controls [38]. This imaging technique holds promise for aiding clinical practice in identifying patients who may benefit from anti-inflammatory therapy. However, its current practical challenges and high costs may limit its widespread clinical application.

Serological biomarkers represent a more accessible and cost-effective alternative for assessing neuroinflammation. However, research in this domain remains limited in the context of lumbar radiculopathy. One study investigated the predictive value of high-sensitivity C-reactive protein (hsCRP) for TEI response in patients with chronic unilateral radiculopathy secondary to spinal stenosis [39]. Patients were stratified based on pre-treatment hsCRP levels, yet no significant differences in outcomes were observed, suggesting that this inflammatory marker may not be a reliable predictor of TEI effectiveness. A fundamental challenge in utilizing serological biomarkers lies in selecting an appropriate reference group for comparison. While TEI response could serve as a surrogate marker, this approach would not directly elucidate the underlying cellular and molecular mechanisms. More invasive methodologies, such as histological analysis of tissue samples obtained during disc surgery, could provide valuable insights, though they are inherently biased toward patients with severe, intractable radiculopathy who undergo surgical intervention.

An alternative, minimally invasive technique for assessing the inflammatory microenvironment surrounding the nerve root is epidural lavage. This method involves the injection and subsequent withdrawal of a small volume of fluid from the epidural space, allowing for the quantification of inflammatory cytokines and peptides [40]. Golish et al. employed epidural lavage to identify a complex of matrix protein degradation fragments associated with inflammatory cytokines as a potential biomarker for ESI response [41]. Although in a small study population, they demonstrated that patients with this complex demonstrated significantly greater functional improvement following ESI compared to those without it. Hence, this might hold promise as a reliable, minimally invasive technique to predict TEI effectiveness.

Advancing research and clinical implications

Although not unequivocally, the current literature provides evidence for trends suggesting that the effectiveness of TEI is largely independent of the underlying aetiology of lumbar radiculopathy, though it may be more effective in patients with a shorter duration of symptoms. Hence, we argue that in patients with a recent onset of a clinically distinct lumbar radiculopathy, the necessity for MRI examination before TEI may be redundant and this therapy can be offered at an earlier stage. This rationale has been the impetus for the TEIAS trial (Transforaminal Epidural Injection in Acute Sciatica) (**Chapter 5**), which aims to investigate the potential benefits of expedited TEI therapy. At present, the long waiting times for clinical consultations by secondary care specialists (i.e., neurologists, orthopaedics and/or neurosurgeons) and imaging assessments – often extending over several weeks or months – result in substantial delays in

treatment initiation. During this period, patients frequently remain impaired in daily functioning and are managed with opioid analgesics. In case of persistent symptoms, some patients may ultimately opt for surgery in pursuit of rapid symptom relief, although long-term patient outcomes do not differ significantly compared to conservative therapy. Hence, the inadequacy of current guidelines in addressing pain management in the acute phase highlights the need for alternative therapeutic strategies. If expedition of TEI proves effective, it could provide general practitioners with an additional therapeutical option to achieve improved symptom management and physical recovery. Simultaneously, this may reduce opioid consumption, work absenteeism and the likelihood of surgery, thereby saving substantial medical and socioeconomic costs.

The TEIAS trial remains ongoing due to the lower-than-anticipated patient enrolment and will be concluded upon reaching the target sample size. A similar study conducted in a regional hospital in the Netherlands was prematurely terminated due to insufficient patient inclusion [42]. This randomized controlled trial examined the effectiveness of TEI in patients with acute lumbar radicular pain (<8 weeks) secondary to MRI-confirmed disc herniation with nerve root impingement, comparing three treatment arms: transforaminal epidural injection with steroid and a local anaesthetic, an injection with saline and a local anaesthetic, and usual care. The protocol intended an enrolment of a total of 264 patients, but in four years the group was able to recruit 141 participants. The trial demonstrated a statistically significant but clinically marginal improvement in leg pain following TEI with steroids and a local anaesthetic compared to usual care (NRS overall mean difference −0.96 (95% CI −1.83 to -0.09; P = 0.030), with no superiority observed for other outcome measures. Moreover, no statistically significant differences were observed between injections with steroids and those with saline (NRS leg pain overall MD −0.28 (95% CI −1.50 to 0.95; P = 0.659). This may be explained by the possible washout effect of cytokines when injecting saline near the affected nerve root. Consequently, the authors concluded that TEI could not be considered effective. In addition, their cost-effectiveness analysis indicated that TEI was not a viable economic alternative [43]. However, due to the limited sample size – 53% of their intended inclusions – there was a lack of statistical power for one of the primary outcome measures. Furthermore, the results were possibly biased by the high loss to follow-up (up to 30%). Since other studies, as summarized in our meta-analysis (**Chapter 2**), were able to demonstrate a statistically significant superiority of steroids over epidural placebo, further research in this area is deemed appropriate.

Although data from the TEIAS trial will be used for subgroup analyses to identify prognostic determinants for responders and non-responders to TEI, the study is

not explicitly designed for that purpose. To address this gap, the POTEISS cohort study (Prediction of Transforaminal Epidural Injection Success in Sciatica) was initiated to develop a prediction model for patient outcomes after TEI (**Chapter 6**). This study systematically evaluates patient outcome after standardized TEI treatment in patients with MRI-confirmed LDH or spinal stenosis causing unilateral radicular symptoms. It was designed with an adequate sample size to facilitate robust prediction modelling, incorporating a comprehensive set of demographical, clinical and radiological variables. The ongoing POTEISS study holds the potential to enhance clinicians' ability to counsel patients on TEI and its expected therapeutical effects, representing an important step toward evidence-based decision-making and more individualized treatment strategies.

In addition to these two studies, future research endeavours should further investigate inflammatory biomarkers and their association with TEI outcomes. Additionally, MC subtypes and treatment response should be assessed in larger studies, ensuring adequate representation of MC type 1, and investigating the usability of other MRI sequences in detecting ongoing inflammation. Furthermore, histological and serological analyses, in the presence and absence of MC, should be correlated with TEI effectiveness. The ongoing EIMICOR cohort study aims to address these aspects and includes patients who have undergone TEI therapy [44]. In addition, epidural lavage represents a promising technique that warrants further validation, alongside advanced imaging modalities such as PET/MRI, which may offer novel insights into neuroinflammation and treatment response. Collectively, these efforts will refine our understanding of the therapeutic potential of TEI, optimize patient selection criteria, and ultimately improve clinical management strategies for lumbar radiculopathy.

THE DAWN OF ARTIFICIAL INTELLIGENCE IN LUMBAR SPINE RESEARCH

Automating diagnostics for spinal stenosis

As previously stated, a key objective in spine research is the advancement toward more personalized treatment strategies. To this end, it is essential to estimate the likelihood of treatment success given a patient's demographic, clinical and radiological profile. For radiological assessments to contribute meaningfully to predictive models, standardized and reliable grading systems are indispensable. Despite the development of comprehensive grading systems that evaluate LSS as a multifaceted degenerative condition, their clinical implementation has remained limited due to suboptimal inter-reader agreement and weak correlations with baseline symptoms and surgical outcomes.

The widely adopted Schizas classification for central canal stenosis (CSS), consisting of seven grades, has demonstrated variable, predominantly moderate, inter-reader agreement [45-49]. Guen et al. reported substantial to near perfect interobserver agreement for their grading system; however, external validation studies revealed only fair agreement, raising concerns about its generalizability [50, 51]. In contrast, the four-grade Miskin scale for CSS consistently demonstrates substantial inter-reader agreement, as confirmed in our validation study (**Chapter 9**), suggesting that it may serve as a more reliable and practical alternative. Moreover, our validation of the Miskin scale for lateral recess stenosis (LRS) also demonstrated substantial agreement, which, together with CCS, supports a structured framework for intra- and interdisciplinary communication regarding spinal stenosis. This is further reinforced by its development through multidisciplinary consultation between radiologists and spine surgeons.

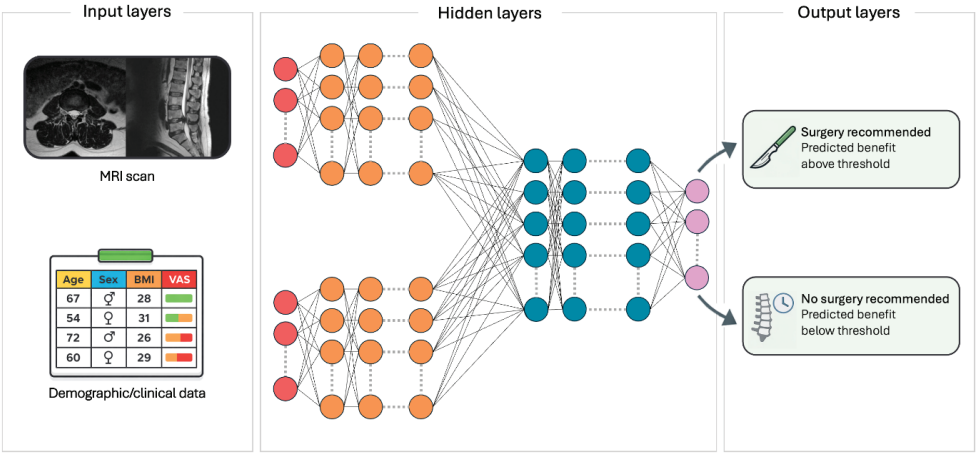
However, the need to eliminate inter-reader variability arises not only from the requirement for precise intra- and interdisciplinary communication but also from its potential contribution to the limited correlation between imaging findings and patient symptoms in previous studies [45-47, 52]. In our validation study of the Miskin scale, higher CCS grades were independently associated with greater improvement in leg pain, back pain and physical disability at 26 and 52 weeks after surgery, underscoring the clinical relevance and applicability of this grading system for preoperative counselling (**Chapter 9**). Nevertheless, clinically meaningful correlations were primarily observed across multiple severity categories, which may be partly attributable to residual inter-reader variability. Recent advances in artificial intelligence (AI), together with the exponential growth of data, offer potential avenues to address this limitation. In particular, AI-based techniques can be applied to automate the classification of radiological imaging, thereby eliminating inter-reader variability. Although machine learning (ML) algorithms have been in development since the 1950s, their recent integration into everyday applications accessible to the general public has sparked widespread interest. Together with the advent of advanced deep learning (DL) models, this has catalysed the use of AI in medical research. In recent years, the number of annual publications addressing segmentation and classification tasks for LSS using conventional ML or DL algorithms has increased substantially, with model performance nearing perfection (**Chapter 10**). However, while numerous novel algorithms have been developed, only few have been externally validated by other research groups using new datasets, limiting assessments of their robustness and generalizability. Additionally, the heterogeneous use of performance metrics presents a challenge in comparing model outcomes across studies. This underscores the need for a standardized set of evaluation metrics that should be consistently reported to ensure compa-

rability and reproducibility. Beyond eliminating inter-reader variability, AI-driven models have demonstrated the potential to significantly reduce workload and reporting time. Studies have shown that AI-assisted radiologists can reduce MRI reporting times by 56 seconds up to 14.1 minutes per study compared to non-AI-assisted radiologists [53, 54]. This efficiency gain represents a significant opportunity for clinicians to reallocate valuable time toward other critical tasks.

An innovative approach to spinal stenosis surgery outcome

Although existing grading scales for LSS severity incorporate multiple items, they may inadvertently omit critical information, thereby failing to correlate effectively with patient symptoms or treatment outcomes. Deep learning models in particular, hold great promise for revolutionizing such imaging analysis tasks. A single axial lumbar spine MRI slice contains 65,536 pixels on average assuming a 256x256 matrix – each pixel representing a specific grey-level intensity. This accumulation of data points results in a highly complex image, making it challenging for human observers to discern all relevant patterns. In contrast, DL algorithms are particularly well-suited for handling high-dimensional data as they can analyse vast numbers of data points simultaneously and identify intricate patterns that might be overlooked by the human eye. This capability was demonstrated in a study on the cervical spine, where a convolutional neural network (CNN) was used to predict clinical success following disc surgery from radiographs [55]. Intriguingly, the CNN's saliency maps revealed that the facet joints – a structure not typically emphasized by neurosurgeons when evaluating these patients for surgical candidacy – had a significant influence on its predictions. Given that current MRI-based degeneration scales for LSS provide limited guidance for surgical decision-making, applying a similar deep learning approach to lumbar spine MRI could uncover imaging features that may be relevant for prognosis but are not readily recognized by clinicians and enhance patient selection for surgery [56-58]. To explore this, we initiated a project to develop a model that uses T2-weighted MRI scans of the stenotic level of operated patients in combination with demographic and clinical data to predict whether a patient will benefit from spinal stenosis surgery. Model training utilized the patient-reported Zurich Claudication Questionnaire, which assesses pain, disability, and postoperative satisfaction at baseline and 26 weeks after surgery [59]. While such high-tech approaches are theoretically compelling, several practical challenges complicate their development. Data acquisition, for instance, is severely constrained by privacy regulations, limiting the size of available datasets. In addition, recruiting skilled engineers to commit to hospital-based research projects is difficult. Consequently, these constraints initially moderated the pace of progress, underscoring that developing a DL

model trained on multimodal data is a complex, time-intensive process that cannot be achieved quickly. Nevertheless, our first model, although not outperforming clinicians, yielded valuable insights that are guiding ongoing refinement and optimization.



Schematic view of a deep learning model for predicting surgical success for LSS patients

The road to AI-driven spinal care

Automating processes will be key in the near future to build a resilient and sustainable healthcare system. As healthcare systems worldwide face increasing pressure due to aging populations, rising costs, and workforce shortages, automation presents an opportunity to enhance efficiency, reduce administrative burdens, and improve patient outcomes. Reports have suggested that up to 35% of healthcare tasks could be automated, allowing medical professionals to dedicate more time to direct patient care [60]. While AI holds immense potential, its integration into clinical decision-making raises concerns about transparency, accountability, and patient trust. At the same time, translating AI models from research to routine neurosurgical practice requires overcoming significant barriers related to data availability, regulatory approval, and clinical validation. On the road to implementation and acceptance, addressing these challenges is crucial to ensuring that AI-driven advancements are both ethically sound and practically feasible.

Ethical frontiers

The integration of AI into medical decision-making presents a range of ethical challenges. A key concern is algorithmic bias, which arises when AI models reflect and potentially amplify existing disparities in healthcare data. Studies have shown that AI systems trained on unrepresentative datasets may perform less

accurately for underrepresented demographic groups, leading to inequitable patient outcomes [61, 62]. Addressing this bias requires diverse, high-quality training data and rigorous validation across different patient populations. Another critical issue is patient consent and transparency. Unlike traditional clinical decision-support tools, DL models function as ''black boxes,'' meaning their decision-making processes are often uninterpretable, even to their developers [63]. Although techniques, such as saliency maps or 'heat maps', have been developed to improve interpretability, they provide only limited insight into the underlying decision-making process and do not fully resolve the black-box nature of deep learning models. This lack of explainability complicates informed consent, as patients and clinicians may struggle to understand how an AI system arrives at its recommendations.

Accountability for AI-driven decisions further complicates matters. If an AI system makes an erroneous diagnosis or treatment recommendation, determining liability—whether it falls on the clinician, the hospital, or the developers of the AI system—remains a legal and ethical grey area [64]. Furthermore, the balance between automation and human oversight is crucial in ensuring AI augments rather than replaces clinical judgment. Over-reliance on AI could lead to deskilling of healthcare professionals, while excessive scepticism may hinder the adoption of beneficial algorithms [65]. Additionally, data privacy and security pose significant ethical challenges, as AI systems require vast amounts of sensitive patient data for training. Robust encryption, federated learning, and strict governance frameworks are necessary to prevent data misuse and unauthorized access [66].

From digital bench to bedside: challenges in AI implementation

The development and implementation of AI-driven algorithms into clinical practice face not only ethical concerns but also several practical challenges. A fundamental requirement for developing robust AI models is access to large, diverse, and high-quality datasets for effective training. However, the collection and sharing of such (anonymous) data is often impeded by privacy concerns, regulatory constraints, and the fragmented nature of healthcare data across institutions and regions. Legislation, such as the General Data Protection Regulation (GDPR) in Europe, adds complexity by imposing stringent data protection measures, including extensive data sharing agreements (DSA). While these measures are essential for safeguarding patient privacy, they strongly delay data exchange across departments, hospitals, and even countries, thereby slowing AI model development.

Once an AI model has been trained, extensive validation is required before it can be considered for clinical use. This requires rigorous assessment through clinical studies mimicking real-world practice to ensure it improves patient outcomes and does not introduce unintended harm or bias. The absence of standardized reporting metrics for AI performance further complicates this process, making it difficult to compare different models. Establishing a minimal set of outcome measures, as we recommended (**Chapter 10**), would facilitate more transparent comparisons and aid clinicians in selecting the most effective AI tools. Furthermore, the "black box" nature of AI models, while primarily an ethical issue, may also present practical challenges. The lack of explainability can hinder AI adoption, as both healthcare providers and patients may be reluctant to trust a system whose reasoning remains unclear. Furthermore, this lack of transparency may lead to hesitance among clinicians, who may prefer maintaining direct control over decision-making, unless they can interpret and justify the model's recommendations. The understanding and interpretation of AI recommendations hinges on the clinician's digital literacy, underscoring the need for integration of AI training into medical education to ensure clinicians are equipped with the necessary skills to effectively utilize AI tools, interpret their outputs, and understand their limitations [67].

Beyond model performance, successful implementation also depends on seamless integration into existing healthcare workflows. AI tools must complement, rather than disrupt, clinical processes, necessitating interoperability with electronic health record systems (HER) and intuitive user interfaces that align with clinicians' decision-making practices. Failure to achieve this alignment may lead to poor adoption rates, even if a model demonstrates clinical efficacy. Currently, this process is complex and time-consuming, requiring collaboration with each individual EHR software supplier to develop tailored integration solutions.

Once an AI model has been developed, validated, and optimized for clinical workflows, it must undergo regulatory approval before it can be deployed in practice. The European Union has introduced stringent standards that AI-driven technologies have to meet under the "Ethics Guidelines for Trustworthy AI", ensuring compliance with principles of transparency, accountability, and fairness, resulting in lengthy and resource-intensive approval processes [68].

Finally, financial constraints present a significant barrier to AI adoption in healthcare. The implementation of AI technologies requires substantial financial investments, yet many healthcare institutions face challenges to allocate sufficient funds for AI integration. Hospitals with limited financial resources often

find it difficult to attract and retain skilled computer engineers, whose expertise is essential for developing and implementing AI models in collaboration with clinicians. Additionally, commercially developed AI solutions can be prohibitively expensive, forcing hospitals to prioritize their investments and potentially limiting the breadth of AI adoption in clinical practice. To justify these investments, it is essential to demonstrate the cost-effectiveness of AI interventions through economic evaluations alongside clinical assessments. Hence, hospitals should be supported by governmental funding to facilitate broader adoption and equitable access to these technologies.

In the context of the aforementioned practical and ethical challenges, the deep learning model we aim to develop to predict surgical success for LSS represents an important first step toward AI-assisted patient selection in spinal surgery. Following model development, clinical translation will require further validation, refinement, and careful navigation of these challenges. Nonetheless, the insights generated thus far support the transformative potential of AI and its capacity to contribute to more precise and data-driven decision-making in spinal care.

REFERENCES

[1] Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, et al. Surgery versus prolonged conservative treatment for sciatica. N Engl J Med. 2007;356(22):2245–56.

[2] van den Hout WB, Peul WC, Koes BW, Brand R, Kievit J, Thomeer RT. Prolonged conservative care versus early surgery in patients with sciatica from lumbar disc herniation: cost utility analysis alongside a randomised controlled trial. Bmj. 2008;336(7657):1351–4.

[3] Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. Cochrane Database Syst Rev. 2016;2016(1):Cd010264.

[4] Costa F, Alves OL, Anania CD, Zileli M, Fornari M. Decompressive Surgery for Lumbar Spinal Stenosis: WFNS Spine Committee Recommendations. World Neurosurg X. 2020;7:100076.

[5] Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976). 2010;35(14):1329–38.

[6] Karppinen J, Malmivaara A, Tervonen O, Pääkkö E, Kurunlahti M, Syrjälä P, et al. Severity of symptoms and signs in relation to magnetic resonance imaging findings among sciatic patients. Spine (Phila Pa 1976). 2001;26(7):E149–54.

[7] el Barzouhi A, Vleggeert-Lankamp CL, Lycklama à Nijeholt GJ, Van der Kallen BF, van den Hout WB, Jacobs WC, et al. Magnetic resonance imaging in follow-up assessment of sciatica. N Engl J Med. 2013;368(11):999–1007.

[8] Athiviraham A, Yen D, Scott C, Soboleski D. Clinical correlation of radiological spinal stenosis after standardization for vertebral body size. Clin Radiol. 2007;62(8):776–80.

[9] Sirvanci M, Bhatia M, Ganiyusufoglu KA, Duran C, Tezer M, Ozturk C, et al. Degenerative lumbar spinal stenosis: correlation with Oswestry Disability Index and MR imaging. Eur Spine J. 2008;17(5):679–85.

[10] van Rijn JC, Klemetso N, Reitsma JB, Majoie CB, Hulsmans FJ, Peul WC, et al. Symptomatic and asymptomatic abnormalities in patients with lumbosacral radicular syndrome: Clinical examination compared with MRI. Clin Neurol Neurosurg. 2006;108(6):553–7.

[11] Boden SD, Davis DO, Dina TS, Patronas NJ, Wiesel SW. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. J Bone Joint Surg Am. 1990;72(3):403–8.

[12] Boos N, Semmer N, Elfering A, Schade V, Gal I, Zanetti M, et al. Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. Spine (Phila Pa 1976). 2000;25(12):1484–92.

[13] Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med. 1994;331(2):69–73.

[14] Takahashi K, Shima I, Porter RW. Nerve root pressure in lumbar disc herniation. Spine (Phila Pa 1976). 1999;24(19):2003–6.

[15] Lindahl O, Rexed B. Histologic changes in spinal nerve roots of operated cases of sciatica. Acta Orthop Scand. 1951;20(3):215–25.
[16] Kato F, Mimatsu K, Kawakami N, Iwata H, Miura T. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. Spine (Phila Pa 1976). 1992;17(8):934–9.
[17] Sawin PD, Traynelis VC, Rich G, Smith BA, Maves TJ, Follett KA, et al. Chymopapain-induced reduction of proinflammatory phospholipase A2 activity and amelioration of neuropathic behavioral changes in an in vivo model of acute sciatica. J Neurosurg. 1997;86(6):998–1006.
[18] Saal JS, Franson RC, Dobrow R, Saal JA, White AH, Goldthwaite N. High levels of inflammatory phospholipase A2 activity in lumbar disc herniations. Spine (Phila Pa 1976). 1990;15(7):674–8.
[19] Franson RC, Saal JS, Saal JA. Human disc phospholipase A2 is inflammatory. Spine (Phila Pa 1976). 1992;17(6 Suppl):S129–32.
[20] Nygaard OP, Mellgren SI, Osterud B. The inflammatory properties of contained and noncontained lumbar disc herniation. Spine (Phila Pa 1976). 1997;22(21):2484–8.
[21] Takahashi H, Suguro T, Okazima Y, Motegi M, Okada Y, Kakiuchi T. Inflammatory cytokines in the herniated disc of the lumbar spine. Spine (Phila Pa 1976). 1996;21(2):218–24.
[22] Burke JG, Watson RW, McCormack D, Dowling FE, Walsh MG, Fitzpatrick JM. Intervertebral discs which cause low back pain secrete high levels of proinflammatory mediators. J Bone Joint Surg Br. 2002;84(2):196–201.
[23] McLain RF, Kapural L, Mekhail NA. Epidural steroid therapy for back and leg pain: mechanisms of action and efficacy. Spine J. 2005;5(2):191–201.
[24] Brisby H, Balagué F, Schafer D, Sheikhzadeh A, Lekman A, Nordin M, et al. Glycosphingolipid antibodies in serum in patients with sciatica. Spine (Phila Pa 1976). 2002;27(4):380–6.
[25] Brisby H, Olmarker K, Rosengren L, Cederlund CG, Rydevik B. Markers of nerve tissue injury in the cerebrospinal fluid in patients with lumbar disc herniation and sciatica. Spine (Phila Pa 1976). 1999;24(8):742–6.
[26] Collighan N, Gupta S. Epidural steroids. Continuing Education in Anaesthesia Critical Care & Pain. 2009;10(1):1–5.
[27] Baqai A, Bal R. The mechanism of action and side effects of epidural steroids. Techniques in Regional Anesthesia and Pain Management. 2009;13(4):205–11.
[28] Tak HJ, Jones R, Cho YW, Kim EH, Ahn SH. Clinical evaluation of transforaminal epidural steroid injection in patients with gadolinium enhancing spinal nerves associated with disc herniation. Pain Physician. 2015;18(2):E177–85.
[29] Djuric N, Yang X, Barzouhi AE, Ostelo R, Duinen SGV, Nijeholt G, et al. Gadolinium Enhancement Is Not Associated With Disc Inflammation in Patients With Sciatica. Spine (Phila Pa 1976). 2019;44(12):E742–e8.
[30] Dudli S, Sing DC, Hu SS, Berven SH, Burch S, Deviren V, et al. ISSLS PRIZE IN BASIC SCIENCE 2017: Intervertebral disc/bone marrow cross-talk with Modic changes. Eur Spine J. 2017;26(5):1362–73.
[31] Modic MT, Masaryk TJ, Ross JS, Carter JR. Imaging of degenerative disk disease. Radiology. 1988;168(1):177–86.

[32] Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology. 1988;166(1 Pt 1):193–9.
[33] Djuric N, Lafeber GCM, Li W, van Duinen SG, Vleggeert-Lankamp CLA. Exploring macrophage differentiation and its relation to Modic changes in human herniated disc tissue. Brain Spine. 2022;2:101698.
[34] Tu GW, Shi Y, Zheng YJ, Ju MJ, He HY, Ma GG, et al. Glucocorticoid attenuates acute lung injury through induction of type 2 macrophage. J Transl Med. 2017;15(1):181.
[35] Kim YH, Kim HJ, Seo J, Chai JW, Song HG, Choi YH, et al. Spinal nerve signal intensity on Dixon T2-weighted water-only sequence: an important outcome predictor after lumbar transforaminal epidural injection. Eur Radiol. 2021;31(12):9459–67.
[36] Albrecht DS, Ahmed SU, Kettner NW, Borra RJH, Cohen-Adad J, Deng H, et al. Neuroinflammation of the spinal cord and nerve roots in chronic radicular pain patients. Pain. 2018;159(5):968–77.
[37] Lutke Schipholt IJ, Scholten-Peeters G, Koop MA, Lammerstma AA, Coppieters MW, van de Giessen E, et al. Quantification of neuroinflammation in spinal cord and neuroforamina in patients with painful cervical radiculopathy using [11C]-DPA713 PET-CT. Submitted for publication.
[38] Lutke Schipholt IJ, Koop MA, Coppieters MW, van de Giessen EM, Lammerstma AA, Ter Meulen BC, et al. Neuroinflammation at the Neuroforamina and Spinal Cord in Patients with Painful Cervical Radiculopathy and Pain-Free Participants: An [(11)C]DPA713 PET/CT Proof-of-Concept Study. J Clin Med. 2025;14(7).
[39] Park CH, Lee SH. Prognostic usefulness of high sensitivity C-reactive protein for transforaminal epidural steroid injection in patients with radicular pain. Pain Med. 2011;12(2):219–23.
[40] Scuderi GJ, Brusovanik GV, Anderson DG, Dunham CJ, Vaccaro AR, Demeo RF, et al. Cytokine assay of the epidural space lavage in patients with lumbar intervertebral disk herniation and radiculopathy. J Spinal Disord Tech. 2006;19(4):266–9.
[41] Golish SR, Hanna LS, Bowser RP, Montesano PX, Carragee EJ, Scuderi GJ. Outcome of lumbar epidural steroid injection is predicted by assay of a complex of fibronectin and aggrecan from epidural lavage. Spine (Phila Pa 1976). 2011;36(18):1464–9.
[42] Ter Meulen BC, van Dongen JM, Maas E, van de Vegt MH, Haumann J, Weinstein HC, et al. Effect of Transforaminal Epidural Corticosteroid Injections in Acute Sciatica: A Randomized Controlled Trial. Clin J Pain. 2023;39(12):654–62.
[43] Ter Meulen BC, Maas ET, van der Vegt R, Haumann J, Weinstein HC, Ostelo R, et al. Cost-effectiveness of Transforaminal epidural steroid injections for patients with ACUTE sciatica: a randomized controlled trial. BMC Musculoskelet Disord. 2024;25(1):247.
[44] Djuric N, Lafeber G, van Duinen SG, Bernards S, Peul WC, Vleggeert-Lankamp CLA. Study protocol: effect of infection, Modic and inflammation on clinical outcomes in surgery for radiculopathy (EIMICOR). BMC Neurol. 2021;21(1):379.
[45] Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine (Phila Pa 1976). 2010;35(21):1919–24.
[46] Moojen WA, Schenck CD, Lycklama À Nijeholt GJ, Jacobs WCH, Van der Kallen BF, Arts MP, et al. Preoperative MRI in Patients With Intermittent Neuro-

genic Claudication: Relevance for Diagnosis and Prognosis. Spine (Phila Pa 1976). 2018;43(5):348–55.
[47] Minetama M, Kawakami M, Teraguchi M, Matsuo S, Enyo Y, Nakagawa M, et al. MRI grading of spinal stenosis is not associated with the severity of low back pain in patients with lumbar spinal stenosis. BMC Musculoskelet Disord. 2022;23(1):857.
[48] Yang J, Xiong Y, Hu Y, Huang M, Zhang L, Pu X, et al. The reliability, correlation with clinical symptoms and surgical outcomes of dural sac cross-sectional area, nerve root sedimentation sign and morphological grade for lumbar spinal stenosis. BMC Musculoskelet Disord. 2023;24(1):225.
[49] Banitalebi H, Espeland A, Anvar M, Hermansen E, Hellum C, Brox JI, et al. Reliability of preoperative MRI findings in patients with lumbar spinal stenosis. BMC Musculoskelet Disord. 2022;23(1):51.
[50] Guen YL, Joon WL, Hee SC, Kyoung-Jin O, Heung SK. A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method. Skeletal Radiology. 2011;40(8):1033–9.
[51] Al-Tameemi HN, Al-Essawi S, Shukri M, Naji FK. Using Magnetic Resonance Myelography to Improve Interobserver Agreement in the Evaluation of Lumbar Spinal Canal Stenosis and Root Compression. Asian Spine J. 2017;11(2):198–203.
[52] Pili M, Tobing SDAL. Functional outcome analysis of lumbar canal stenosis patients post decompression and posterior stabilization with stenosis grading using magnetic resonance imaging. Journal of Physics: Conference Series. 2017;884(1):012146.
[53] Lim DSW, Makmur A, Zhu L, Zhang W, Cheng AJL, Sia DSY, et al. Improved Productivity Using Deep Learning-assisted Reporting for Lumbar Spine MRI. Radiology. 2022;305(1):160–6.
[54] Yang YXC, Yee SY, Tan TSE, Koh KKN, Goh AGW, Venugopal VK, et al. An artificial intelligence boost to MRI lumbar spine reporting. Eur J Radiol. 2024;179:111636.
[55] Goedmakers C, Pereboom L, de Leeuw den Bouter M, Remis R, Staring M, Vleggeert-Lankamp C. Deep Learning on Preoperative Radiographs for Clinical Success Prediction after Surgery for Cervical Degenerative Disease. Brain and Spine. 2023;3:101842.
[56] Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, et al. Deep Learning in Medical Imaging. Neurospine. 2019;16(4):657–68.
[57] Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med. 2021;4(1):65.
[58] Liu X, Gao K, Liu B, Pan C, Liang K, Yan L, et al. Advances in Deep Learning-Based Medical Image Analysis. Health Data Sci. 2021;2021:8786793.
[59] Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine (Phila Pa 1976). 1996;21(7):796–803.
[60] Spatharou A, Hieronimus S, Jenkins J. Transforming healthcare with AI: The impact on the workforce and organizations. McKinsey & Company; 2020.
[61] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
[62] Colacci M, Huang YQ, Postill G, Zhelnov P, Fennelly O, Verma A, et al. Sociodemographic bias in clinical machine learning models: a scoping review of algorithmic bias instances and mechanisms. J Clin Epidemiol. 2025;178:111606.

[63] Samek W, Wiegand T, Müller K-R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. arXivorg. 2017.
[64] Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare. 2020:295–336.
[65] Choudhury A, Chaudhry Z. Large Language Models and User Trust: Consequence of Self-Referential Learning Loop and the Deskilling of Health Care Professionals. J Med Internet Res. 2024;26:e56764.
[66] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digital Medicine. 2020;3(1):119.
[67] Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ. 2019;5(2):e16048.
[68] European Commission. Ethics Guidelines for Trustworthy AI. 2019.