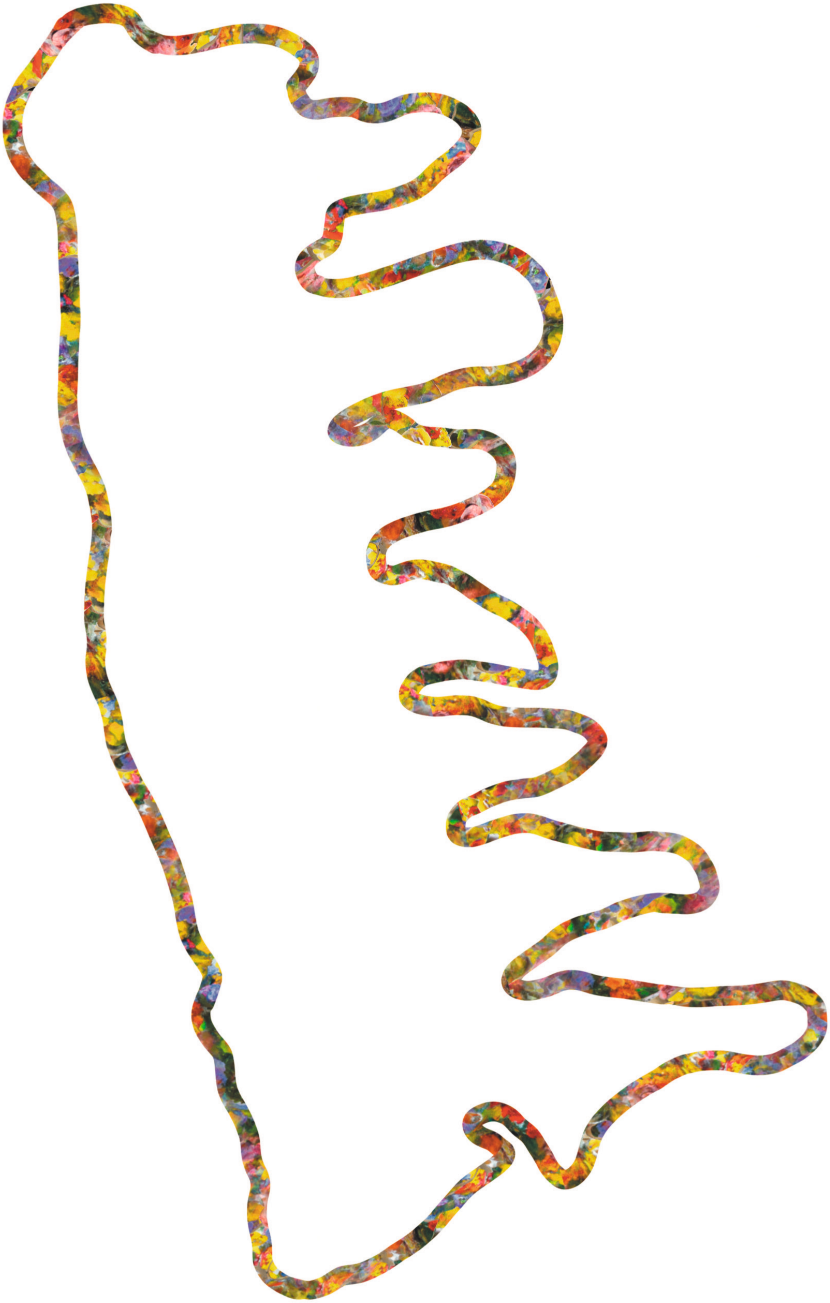
Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning

Goedmakers, C.M.W.



Appendices

Summary
Samenvatting
Acknowledgements
Curriculum Vitae
List of publications

Summary

The progressive degenerative changes in the cervical spine can result in neck pain, reduced functioning, gradual loss of normal structure, and conditions such as sagittal malalignment, radiculopathy, and myelopathy. Anterior cervical discectomy and fusion (ACDF) is the standard surgical treatment for degenerative cervical pathology, but around 25% of patients report low satisfaction after surgery, related to persisting disability. The difference is hypothesized to be caused by different factors including; type of surgical procedure, baseline degenerative patterns in the spine and mental health status of the patient. Identifying the underlying pathophysiological process causing the persisting disability after surgery would allow for preoperative recognition of the recovery pathway a patient will follow after surgery.

In this thesis, classical statistical methods were applied to clinical data, and Machine Learning was used on medical imaging data, in an effort to determine the factors that contribute to the variation in, clinically important, functional outcomes following anterior cervical discectomy and to explore how artificial intelligence can be utilized to improve the diagnostic and prognostic process.

In literature, placing a prosthesis after anterior discectomy (ACDA) is believed to decrease complaints of Adjacent Segment Disease (ASD). Previously the comparison of fusion (ACDF) to prosthesis (ACDA) in the treatment of cervical radiculopathy was made based on studies including mixed patient populations: radiculopathy patients with and without myelopathy. Aiming to compare the true clinical effect of the prosthesis, in literature, to that of fusion surgery, clinical outcome after ACDA was compared to ACDF in patients exclusively suffering from radiculopathy, and pain and disability scores were comparable in patients after 2 years and not dependent the type of surgical procedure. Additionally, the differences found in the mixed, myelopathy and radiculopathy, patient populations were not clinically relevant. The same was true for the reoperation rates and the incidence of adjacent segment degeneration (ASD).

Defining clinically meaningful outcome is crucial for clinical audits, decision-making and research. In studying criteria for a successful outcome 3 and 12 months after surgery for cervical degenerative radiculopathy on recommended PROMs, a Neck Disability Index (NDI) percentage change score of 35.1, one year after surgery, demonstrated highest ability for distinguishing between a clinically successful and not successful outcome (AUC 0.91), as it correlated well with patients their self-reported satisfaction. For all recommended PROMs discriminative ability was better for the percentage change scores and the follow-up scores, compared to the change scores.

After combining two randomized, double-blinded clinical trials on anterior decompression in cervical radiculopathy it was demonstrated that at two year follow up a clinical advantage for the cervical disc prosthesis is absent, when compared to the golden standard ACDF. In contrast to what is globally hypothesized by many spinal surgeons, superior outcome after ACDA could not even be confirmed when ACDA was compared to ACD, without instrumentation. Additionally, preliminary subgroup analysis could not indicate a certain type of patient that would benefit more from receiving the prosthesis two years after surgery.

However, in post-hoc analysis of randomized controlled trial (RCT) data it was found that patients that were classified as depression or anxiety cases at baseline had statistically significant and clinically relevant higher NDI scores one and two years after surgery. Linear Mixed Models (LMM) enabled the successful analysis of decreased mental health using HADS on a continuous scale and demonstrated that if during follow-up if the NDI improved, so did the Hospital Anxiety and Depression score (HADS). Additionally, an R-shiny application was developed to facilitate an easier-to-interpret, visual communication of the LMMs to patients during a preoperative clinic visit.

As adjacent segment disease is typically reported to occur after longer follow-up, it is essential to study long-term results in patients undergoing anterior discectomy to treat cervical radiculopathy. In the long-term follow-up of the RCT comparing ACDA, ACDF and ACD a persisting absence of clinical superiority was demonstrated for the cervical disc prosthesis five years after surgery. Specifically, clinically relevant adjacent level disease was not prevented by implanting a prosthesis. Single level ACD without implanting an intervertebral device provided worse clinical outcome, which was hypothesized to be caused by delayed fusion

There is immense potential for healthcare to benefit from the two largest technological innovations of the twenty-first century: big data and artificial intelligence (AI). Machine Learning, a fundamental driver of current-generation AI, feeds on big data to identify underlying patterns to improve its model performance on predefined tasks. With an explosion of healthcare data ranging from basic science, clinical, numerical, language-based, and imaging data to vast amounts of administrative and economic data, artificially intelligent computer programs are well positioned to harvest, manage, analyze, and interpret this data to optimize healthcare delivery. However, the rise of big data and related technologies has led to an increased reliance on computer algorithms and a decline in direct human involvement and oversight. This shift raises ethical concerns in healthcare, as the machine-driven automation of processes, that were previously guided by human intelligence, creates new challenges. For the ethical application of Machine Learning in healthcare clinicians should balance the principles of beneficence, non-maleficence, justice and respect for autonomy to ensure patient safety, reliable diagnoses, and fair treatment. To guarantee beneficence and non-maleficence, AI needs to be primarily accurate and reliable. The data used to train AI algorithms must be carefully curated to avoid biases that could result in harm or injustice. Additionally, the healthcare provider must maintain patient autonomy by providing necessary information, regarding the underlying technology, for informed decision making.

Automating parts of the radiological image analysis process using Machine Learning could provide more accurate and consistent assessment, eliminating interobserver variability with an increased time efficiency. We found a wide variety in literature of possibilities in Machine Learning methods, depending on the aim of the application and the available modalities. In segmentation models, Deep Learning methods show promising results with the application of (fully automatic) Convolutional Neural Network (CNN) models using X-ray, CT or MR imaging. Regarding cervical spine analysis, the biomechanical properties are most often studied using finite element models.

The application of artificial neural networks and support vector machine models looks promising for other classification purposes.

Consequently, a CNN was developed to identifying patients who will develop ASD, using only preoperative cervical MRI in patients undergoing single-level anterior cervical decompressive surgery. Predicting ASD remains challenging for clinicians. In 344 patients who underwent single-level anterior discectomy and fusion, a deep learning algorithm outperformed two clinical experts at predicting new symptoms at adjacent levels, solely based on the preoperative MRI (95% vs 58% accuracy, respectively). In identifying patients that would develop symptoms, the compression of neural structures appeared to be the most important feature for the deep learning algorithm

A problem that initiated this research is that not all patients benefit from surgical treatment for degenerative cervical disc disease and predicting who will is difficult. Therefore, in the final study included in this thesis, a CNN was developed to predict clinical success one year after surgery for cervical degenerative disease, solely based on preoperative lateral radiographs, and identify which image features are important for the algorithms decision-making.The model showed good discriminative ability in accuracy, sensitivity and AUC. The Grad-CAM heatmaps of this study show a significant influence of the facet joints on the classification process.

The clinical issue that initiated this thesis was the challenge of determining which patients will not benefit from surgical treatment for degenerative cervical disc disease. Therefore, as part of the final study included in this thesis, a CNN was developed to predict clinical success one year after surgery for cervical degenerative disease, based solely on preoperative lateral radiographs. The aim was to identify which image features played a key role in the algorithm's decision-making process. The model demonstrated a high level of discriminatory ability in terms of accuracy, sensitivity and AUC. The Grad-CAM heatmaps of the study revealed substantial influence of the facet joints on the classification process, the only synovial-lined, diarthrodial joints of the spine.