



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2024

---

## **Identifying clusters of objective functional impairment in patients with degenerative lumbar spinal disease using unsupervised learning**

Staatjes, Victor E ; Klukowska, Anita M ; Stumpo, Vittorio ; Vandertop, W Peter ; Schröder, Marc L

DOI: <https://doi.org/10.1007/s00586-023-08070-z>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-251945>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Staatjes, Victor E; Klukowska, Anita M; Stumpo, Vittorio; Vandertop, W Peter; Schröder, Marc L (2024). Identifying clusters of objective functional impairment in patients with degenerative lumbar spinal disease using unsupervised learning. *European Spine Journal*, 33(4):1320-1331.

DOI: <https://doi.org/10.1007/s00586-023-08070-z>



# Identifying clusters of objective functional impairment in patients with degenerative lumbar spinal disease using unsupervised learning

Victor E. Staartjes<sup>1,2,3</sup> · Anita M. Klukowska<sup>2,3,4</sup> · Vittorio Stumpo<sup>1</sup> · W. Peter Vandertop<sup>2</sup> · Marc L. Schröder<sup>3</sup>

Received: 30 April 2023 / Revised: 22 October 2023 / Accepted: 27 November 2023  
© The Author(s) 2023

## Abstract

**Objectives** The five-repetition sit-to-stand (5R-STS) test was designed to capture objective functional impairment (OFI), and thus provides an adjunctive dimension in patient assessment. It is conceivable that there are different subsets of patients with OFI and degenerative lumbar disease. We aim to identify clusters of objectively functionally impaired individuals based on 5R-STS and unsupervised machine learning (ML).

**Methods** Data from two prospective cohort studies on patients with surgery for degenerative lumbar disease and 5R-STS times of  $\geq 10.5$  s—indicating presence of OFI. *K*-means clustering—an unsupervised ML algorithm—was applied to identify clusters of OFI. Cluster hallmarks were then identified using descriptive and inferential statistical analyses.

**Results** We included 173 patients (mean age [standard deviation]: 46.7 [12.7] years, 45% male) and identified three types of OFI. OFI Type 1 (57 pts., 32.9%), Type 2 (81 pts., 46.8%), and Type 3 (35 pts., 20.2%) exhibited mean 5R-STS test times of 14.0 (3.2), 14.5 (3.3), and 27.1 (4.4) seconds, respectively. The grades of OFI according to the validated baseline severity stratification of the 5R-STS increased significantly with each OFI type, as did extreme anxiety and depression symptoms, issues with mobility and daily activities. Types 1 and 2 are characterized by mild to moderate OFI—with female gender, lower body mass index, and less smokers as Type I hallmarks.

**Conclusions** Unsupervised learning techniques identified three distinct clusters of patients with OFI that may represent a more holistic clinical classification of patients with OFI than test-time stratifications alone, by accounting for individual patient characteristics.

**Keywords** Objective functional testing · Functional impairment · Diagnostics · Classification · Machine learning · Unsupervised machine learning

## Introduction

The clinical assessment of patients suffering from back and leg pain due to lumbar degenerative disease has recently been supplemented by tests for objective functional impairment (OFI) [1–6]. Tests that have been well validated include the timed-up-and-go, 6-min-walk, and five-repetition sit-to-stand (5R-STS) tests [1, 7, 8]. In addition to clinical examination and questionnaire measures for pain and subjective functional impairment, these tests have been shown to be robust to mental status as a confounder and add the ability to capture deficits and complications, such as foot drop or limping [2, 9]. Patients also prefer objective tests over a battery of questionnaires to assess functional impairment [10, 11]. When applied together with questionnaires for pain severity, subjective functional impairment and health-related quality of life, these tests provide a holistic capture

✉ Victor E. Staartjes  
victoregon.staartjes@usz.ch

<sup>1</sup> Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Department of Neurosurgery, Clinical Neuroscience Center, University Hospital Zurich, University of Zurich, Frauenklinikstrasse 10, 8091 Zurich, Switzerland

<sup>2</sup> Amsterdam UMC, Vrije Universiteit Amsterdam, Neurosurgery, Amsterdam Movement Sciences, Amsterdam, The Netherlands

<sup>3</sup> Department of Neurosurgery, Bergman Clinics, Amsterdam, The Netherlands

<sup>4</sup> Department of Surgery, Royal Derby Hospital, Derby, UK

of a patient's health state for scientific and clinical purposes [12–14].

A 5R-STS test time of 10.5 s or greater has been shown to correspond to a diagnosis of OFI based on normative data [1]. Baseline severity stratifications have also been constructed, specifying cut-offs for mild, moderate, and severe OFI [15, 16]. However, these cut-offs assume a similar performance among normative populations across all sociodemographic groups. In reality, older patients, those with higher BMI, active smokers, taller patients, and many other groups do worse on the 5R-STS. Cut-offs should be calculated from normative data across all of these groups, but the cut-offs should be flexible and adjustable to an individual's characteristics.

Achieving such “personalized” cut-offs for OFI can be achieved by calculating cut-offs for specific populations, e.g. cut-offs for  $>/< 65$  years of age and for male and female individuals [1]. However, this would result in a great number of different cut-offs that would be hard to implement in clinical practice. In the era of “personalized/precision medicine”, a more elegant option is to predict the expected upper limit of normal (ULN) for individual patients, based on their sociodemographic characteristics, in order to diagnose OFI [17, 18]. This works well for single cut-offs, e.g. for the binary presence or absence of OFI, based on normative data, but is not suitable for identifying mild, moderate, and severe impairment. These subgroups should instead be defined according to real-world data of patients with established OFI and should reflect specific hallmarks of these subgroups. For example, classifying severity only according to 5R-STS results (e.g. based on test time cut-offs for mild, moderate, or severe disease) would not take into account inter-individual differences among patients. Unsupervised machine learning techniques, such as clustering, are well suited for identifying clusters of observations that exhibit high similarity, without providing labels (e.g. “mild”, “moderate”, “severe”) [19–22]. Clusters defined by a machine learning algorithm would not be based on disease-specific parameters and could then be used to classify new patients into relevant subsets that may also exhibit differences in treatment response. We aimed to identify clusters of OFI in objectively functionally impaired individuals based on 5R-STS and unsupervised machine learning methods.

## Materials and methods

### Study design

Pooled data from two prospective studies were used: ClinicalTrials.gov Identifiers: NCT03303300 and NCT03321357 [1, 23]. Both studies were approved by the local institutional review board (Medical Research Ethics Committees

United, Registration Numbers: W17.107 and W17.134) and were conducted according to the Declaration of Helsinki. Informed consent was obtained from all participants.

Patients scheduled for lumbar spine surgery for degenerative disease at a Dutch specialized short stay spine clinic were included between October 2017 and June 2018 and were assessed during outpatient consultations. Participating patients completed a variety of questionnaires, as well as the 5R-STS test. The pooled data from both studies were used to train an unsupervised machine learning model to automatically identify clusters of OFI. Subsequently, we compared the identified clusters to identify their hallmarks for further interpretation.

### Inclusion and exclusion criteria

Inclusion criteria were the presence of lumbar disc herniation, lumbar spinal stenosis, spondylolisthesis, or discogenic chronic low back pain. Patients with synovial facet cysts causing radiculopathy, hip or knee prosthetics, and those requiring walking aides were excluded to eliminate these confounders. We also excluded all healthy volunteers, who were recruited in the control group. In addition, we excluded all patients without OFI (i.e. a 5R-STS test time of  $< 10.5$  s, as defined by Staartjes et al. [1]) in order to cluster only those patients with established OFI.

### Data collection

The 5R-STS was performed according to the protocol described by Jones et al. [5] and Staartjes et al. [1] If the patient was unable to perform the test in 30 s, or not at all, this was noted and the test score was recorded as 30 s. [1] The baseline severity stratification for the 5R-STS, validated by Klukowska et al. [15], was used. Patients also filled in questionnaires containing baseline sociodemographic data including age, gender, smoking status, body mass index (BMI), prior spine surgery, indication and index level, history of complaints, education, work type and ability, analgesia, symptom satisfaction, as well as numeric rating scales for back and leg pain severity and validated Dutch versions of the Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RMDQ), and EuroQOL-5D-3L (EQ-5D) to capture subjective functional impairment, as well as HRQOL [24–26]. The EQ-5D included its single domains as well as the composite EQ-5D index and the EQ-5D thermometer on current subjective health status [26]. Participants filled out the questionnaires right after initially performing the test during outpatient consultation. For the EQ-5D, it has been established that the mood component of the EQ-5D correlates well with clinical depression [27].

## Analytical methods

Data were reported as mean (standard deviation) for continuous and numbers (percentages) for categorical data. Variables with missingness over 25% were not included in the analysis. When data were assumed to be missing at random (MAR) or missing completely at random (MCAR), imputation was performed using a  $k$ -nearest neighbour (KNN) imputation with  $k = 5$  [28]. Pearson's product-moment correlation was applied to provide an overview of correlations within the dataset—that is, to identify which variables appear to be most highly correlated with 5R-STTS results, and which variables demonstrate multicollinearity. One-way analysis of variance (ANOVA) or Pearson's Chi-Square tests were performed to test for differences among the identified clusters, for continuous and categorical variables, respectively. A  $p \leq 0.05$  on two-sided tests was considered significant. Analyses were carried out using R version 4.0.3 (The R Foundation for Statistical Computing, Vienna, Austria) [29].

We chose  $k$ -means clustering to carry out unsupervised clustering of patients with OFI. The optimal number of clusters was chosen using the “elbow method” based on within-cluster sum of squares. Briefly, this method identifies the number of  $k$  clusters from which onwards the increase in similarity of observations within clusters becomes linear. The version of the  $k$ -means clustering algorithm described by Hartigan and Wong was used [30]. Pre-processing included centring and scaling (standardization), as well as one-hot encoding of categorical variables. The algorithm was run for a maximum number of iterations of 1000, with 100 initial configurations. Only the 5R-STTS test time, 5R-STTS baseline severity stratification, patient age, gender, height, weight, BMI, and smoking status were provided as inputs to the model, as sociodemographic variables unspecific to disease, as opposed to e.g. back pain severity or index level. A KNN algorithm with  $k = 5$  was subsequently trained to classify new patients into the corresponding clusters.

## Results

### Patient cohort

We included 173 patients with OFI fulfilling the inclusion criteria. Detailed characteristics are provided in Table 1. Data missingness was 3.5%. Mean age was 46.72 years (12.65), and 78 patients (45.1%) were male. According to the validated baseline severity stratification, 95 patients (54.9%) had mild, 45 (26.0) had moderate, and 33 (19.1%) had severe OFI. A correlation matrix of all variables included in the model is shown in Fig. 1.

**Table 1** Baseline characteristics of the overall patient cohort

Parameter	Value (N= 173)
5R-STTS Test time, mean (SD)	16.88 (6.24)
<i>Functional impairment group, n (%)</i>	
Mild (10.5–15.2 s.)	95 (54.9)
Moderate (15.3–22.0 s.)	45 (26.0)
Severe (> 22.0 s.)	33 (19.1)
Age, mean (SD)	46.72 (12.65)
Male gender, n (%)	78 (45.1)
Height, mean (SD)	175.91 (9.81)
Weight, mean (SD)	79.36 (12.88)
Body mass index, mean (SD)	25.57 (3.32)
<i>Smoking status, n (%)</i>	
Active smoker	56 (32.4)
Ceased	52 (30.1)
Never smoked	65 (37.6)
Prior surgery, n (%)	37 (21.4)
<i>Indication for surgery, n (%)</i>	
Lumbar disc herniation	127 (73.4)
Stenosis	29 (16.8)
Spondylolisthesis	7 (4.0)
Chronic low back pain	10 (5.8)
<i>History of complaints, n (%)</i>	
< 6 wks	7 (4.0)
6 wks.–3 months	29 (16.8)
6 months–1 year	49 (28.3)
> 1 year	88 (50.9)
<i>Index level, n (%)</i>	
L2–L3	6 (3.5)
L3–L4	10 (5.8)
L4–L5	75 (43.4)
L5–S1	82 (47.4)
<i>Highest level of education, n (%)</i>	
Elementary school	3 (1.7)
High school	80 (46.2)
Higher education	84 (48.6)
(Post-)doctoral	6 (3.5)
<i>Type of work, n (%)</i>	
Student	1 (0.6)
Houseworker	7 (4.0)
Employed	106 (61.3)
Self-employed	28 (16.2)
On benefits	7 (4.0)
Retired	14 (8.1)
Jobless	10 (5.8)
<i>Analgesic medication, n (%)</i>	
Not regularly	21 (12.1)
At least weekly	17 (9.8)
Daily	135 (78.0)
<i>Satisfied with current symptoms, n (%)</i>	
Yes	2 (1.2)
Neutral	3 (1.7)

**Table 1** (continued)

Parameter	Value (N=173)
No	168 (97.1)
Ability to work, n (%)	
Fully able	42 (24.3)
Limited	28 (16.2)
Unable	103 (59.5)
<i>EQ-5D mobility, n (%)</i>	
No problems	13 (7.5)
Some problems	140 (80.9)
Confined to bed	20 (11.6)
<i>EQ-5D Selfcare, n (%)</i>	
No problems	83 (48.0)
Some problems	88 (50.9)
Unable	2 (1.2)
<i>EQ-5D daily activities, n (%)</i>	
No problems	8 (4.6)
Some problems	112 (64.7)
Unable	53 (30.6)
<i>EQ-5D Pain, n (%)</i>	
No pain or discomfort	5 (2.9)
Moderate pain or discomfort	54 (31.2)
Extreme pain or discomfort	114 (65.9)
<i>EQ-5D Mood, n (%)</i>	
Not anxious or depressed	101 (58.4)
Moderately anxious or depressed	59 (34.1)
Extremely anxious or depressed	13 (7.5)
EQ-5D Index, mean (SD)	0.33 (0.30)
EQ-5D Thermometer, mean (SD)	46.89 (17.51)
NRS back pain, mean (SD)	6.57 (2.36)
NRS leg pain, mean (SD)	7.57 (1.82)
Oswestry disability index, mean (SD)	48.73 (16.42)
Roland-Morris disability questionnaire, mean (SD)	13.67 (5.05)

*5R-STS* five-repetition sit-to-stand test, *SD* standard deviation, *EQ-5D* EuroQOL five-dimensions questionnaire, *NRS* Numeric rating scale

## Clustering analysis

A plot of within-cluster sum of squares against the number of clusters (Fig. 2) indicated that a number of clusters between 3 and 6 would constitute the optimal  $k$ , as this is the point from which onwards the similarity among observations within the clusters only increases marginally. For the analysis,  $k = 3$  was chosen.

The three identified clusters (Types 1 to 3) contained 57 (32.9%), 81 (46.8%), and 35 (20.2%) patients, respectively. Within-cluster sum of squares values were 209, 363, and 167, respectively. The ratio of between-cluster sum of squares and total sum of squares was 34.1%.

## Cluster hallmarks

### Clustered variables

Table 2 provides an overview of the differences between the three clusters in terms of the variables that was included in the model. The clusters of impairment are illustrated in Fig. 3 for continuous variables and Fig. 4 for categorical variables. In terms of raw test times, Type 1 and Type 2 were comparable with mean test times between 14 and 15 s, while Type 3 demonstrated a mean test time of 27.1 (4.4) seconds. The distribution of mild, moderate, and severe OFI groups according to the validated 5R-STS baseline severity stratification increased steadily from Type 1 to Type 3 [15]. Age was constant across all clusters. When comparing Type 1 and Type 2 OFI, the rate of smokers and males was significantly lower in Type 1, as were mean BMI and body height.

### Unclustered variables

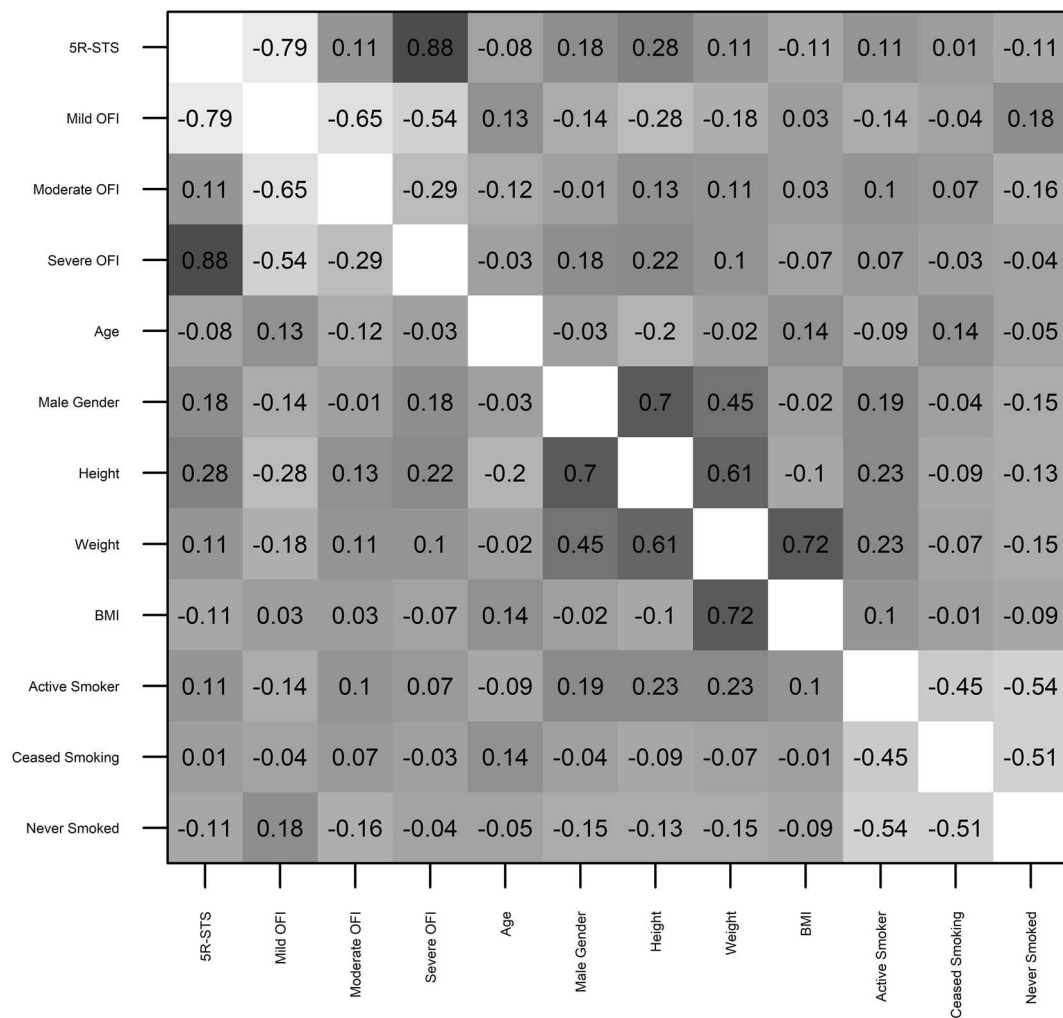
To further characterize types of OFI, those variables not included in the clustering analysis ought to be analysed (Table 3). There were marked differences in all EQ-5D domains, as well as the EQ-5D index and EQ-5D thermometer and the ODI and RMDQ. Specifically, the rate of patients with extreme anxiety and depression increased steadily from 3.5% in Type 1, 7.4% in Type 2, to 14.3% in Type 3, with statistical significance. In addition, mobility and ability to perform activities of daily life (ADL) were reduced in Type 3, with corresponding increases in subjective functional impairment scores (ODI, RMDQ).

The proportion of patients who had undergone prior spine surgery increased steadily from Type 1 with 14.0% to Type 3 with 28.6%, although this progression was not statistically significant. There were no differences in back or leg pain severity in the three identified clusters. Similarly, indications for surgery, history of complaints, index levels, education, work type and ability, analgesic medication use, and satisfaction also remained constant across all three clusters. A qualitative overview of the hallmarks of each type is provided in Table 4.

## Discussion

Three characteristic clusters of patients with OFI were identified through unsupervised analysis. The clusters were termed Types 1, 2 and 3, and roughly correspond to mild, moderate, and severe impairment (Table 4).

Type 1 OFI was present in around a third of patients and was characterized by a relatively rapid performance of the 5R-STS and was only seldomly associated with problems in performing ADL, mobility, and clinical



**Fig. 1** Correlation matrix for 5R-STS test time, baseline severity stratification (BSS), age, gender, height, weight, body mass index (BMI), and smoking status. Pearson's product-moment correlation is

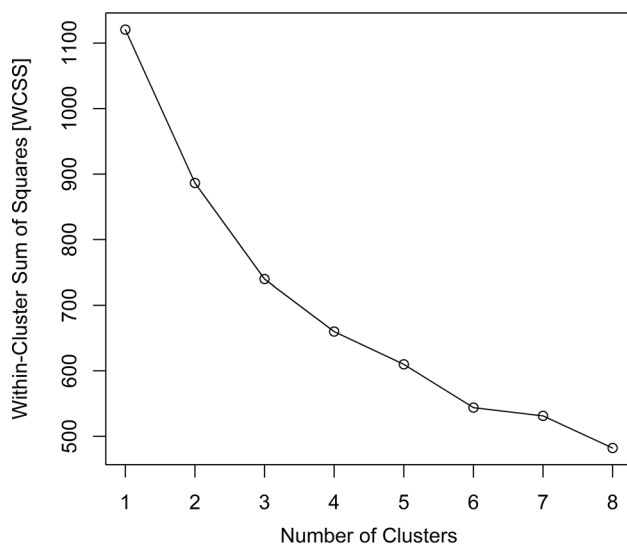
demonstrated. *5R-STS* five-repetition sit-to-stand test, *OFI* objective functional impairment, *BMI* body mass index

depression—indicating mild impairment. This is also supported by the low levels of subjective functional impairment found in these patients. As mentioned in Results section, concerning demographics, the vast majority of Type 1 patients were female nonsmokers, with a low BMI. The female gender also explains the lower average height in this group. It has been argued that female patients have a higher pain tolerance than male patients and are likely to also present later for surgical treatment for degenerative spinal conditions [31–34]. This could partially explain that this largely female group experiences low subjective and OFI. The low incidence of active smoking demonstrably has no effect on 5R-STS performance [1] and for that matter also not on other short-duration objective functional tests [35]. This likely indicates that smoking, while not a significant predictor of 5R-STS performance, was picked up by the clustering algorithm as a confounder associated

with other, possibly psychosocial factors that in turn influence performance.

Type 2 OFI occurred in half of our cohort and was linked with overweight in both genders, although test times were slightly elevated compared to Type 1. This indicates mild impairment, also corroborated by mild subjective functional impairment. As stated in results, the incidence of extreme anxiety and depression symptoms was over twice as high as in Type 1, and statistically significantly. The rate of smokers corresponded to that of our patient cohorts and indeed the Dutch population [36]. In addition, both genders were equally represented in this cluster. Type 2 likely indicates low levels of true functional impairment, but with a higher susceptibility for mood changes due to the mild or moderate impairment that is present.

In contrast to the mild/moderate levels of OFI observed in Types 1 and 2, Type 3 indicated extreme impairment with



**Fig. 2** Plot of within-cluster sum of squares (WCSS) against number of clusters. The number of clusters at which the decrease in WCSS becomes linear ought to be chosen as the number of clusters for k-means clustering based on the “elbow” method. In more detail, the WCSS describes the distance between each observation and the centroid within each cluster, i.e. how well the cluster fits the single observations within it. With an increasing number of clusters, WCSS decreases because clusters become more specific. However, after reaching a certain number of clusters, WCSS starts to decrease much more slowly. This “elbow” point often provides an optimal balance between a low amount of clusters (allowing for sensible interpretation of clusters) which still adequately represent the data

sequelae such as bedriddenness, high subjective functional impairment, mobility issues, and high rates of discomfort. Overall, patients with Type 3 impairment were of average BMI, mostly of male gender, and exhibited a significantly higher rate of active smoking. We also observed a doubling of the rate of extreme depression and anxiety compared to Type 2 and a quadrupling compared to Type 1.

Still, levels of pain were comparable among the three clusters, with the exception of the EQ-5D “pain and discomfort” domain, which also includes discomfort. Back and leg pain severity did not differ among the three clusters, demonstrating that the clusters represent true subgroups of impairment (including objective and subjective impairment), and are not influenced by pain severity as such. This is similar for sociodemographic factors, which could be assumed to influence the perception of impairment, such as level of education, work type, work ability, and age.

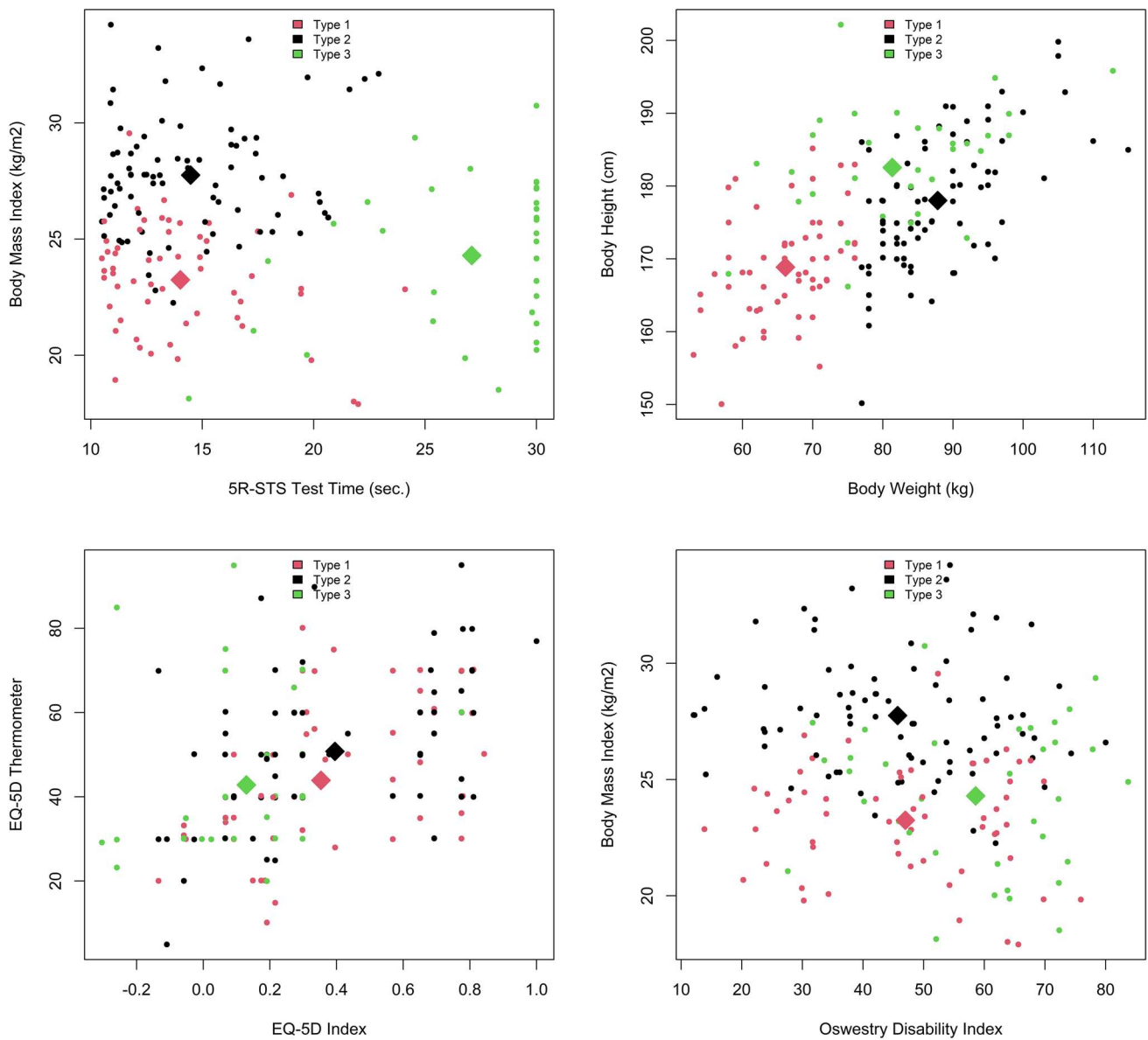
Up to now, grading of OFI was based on a fixed cut-off of 10.5 s on the 5R-STST test, though, realistically, obese and elderly, but otherwise healthy, individuals cannot be expected to perform equally well as younger individuals with a BMI in the normal range [1]. Ideally, an otherwise healthy, but obese, 75-year-old person and a 22-year-old athlete should not have their level of impairment rated by the same static cut-off. As one potential solution, Gautschi et al. [2] calculated a range of cut-offs for patients of certain gender and ages for the timed-up-and-go test, but clinical implementation of a larger amount of cut-offs that need to be remembered is cumbersome. Machine

**Table 2** Comparative analysis of the three types of objective functional impairment identified in the clustering analysis by means of those variables included in the clustering analysis (Clustered Parameters)

Parameter	Type 1 impairment	Type 2 impairment	Type 3 impairment	<i>P</i>
<i>N</i> , (%)	57 (32.9)	81 (46.8)	35 (20.2)	
<i>Clustered parameters</i>				
5R-STST test time, mean (SD)	14.02 (3.23)	14.48 (3.29)	27.09 (4.42)	<0.001*
Age, mean (SD)	46.47 (12.63)	48.55 (12.98)	42.86 (11.27)	0.082
Male gender, <i>n</i> (%)	9 (15.8)	44 (54.3)	25 (71.4)	<0.001*
Height, mean (SD)	168.86 (7.48)	178.01 (8.89)	182.54 (8.35)	<0.001*
Weight, mean (SD)	66.11 (6.30)	87.82 (8.23)	81.34 (12.05)	<0.001*
Body mass index, mean (SD)	23.25 (2.36)	27.75 (2.50)	24.29 (3.14)	<0.001*
<i>Smoking status, n (%)</i>				
Active smoker	9 (15.8)	31 (38.3)	16 (45.7)	
Ceased	19 (33.3)	23 (28.4)	10 (28.6)	
Never smoked	29 (50.9)	27 (33.3)	9 (25.7)	
<i>Functional impairment group, n (%)</i>				
Mild (10.5–15.2 s.)	43 (75.4)	51 (63.0)	1 (2.9)	<0.001*
Moderate (15.3–22.0 s.)	13 (22.8)	28 (34.6)	4 (11.4)	
Severe (> 22.0 s.)	1 (1.8)	2 (2.5)	30 (85.7)	

5R-STST five-repetition sit-to-stand test, *SD* standard deviation

\* $p \leq 0.05$



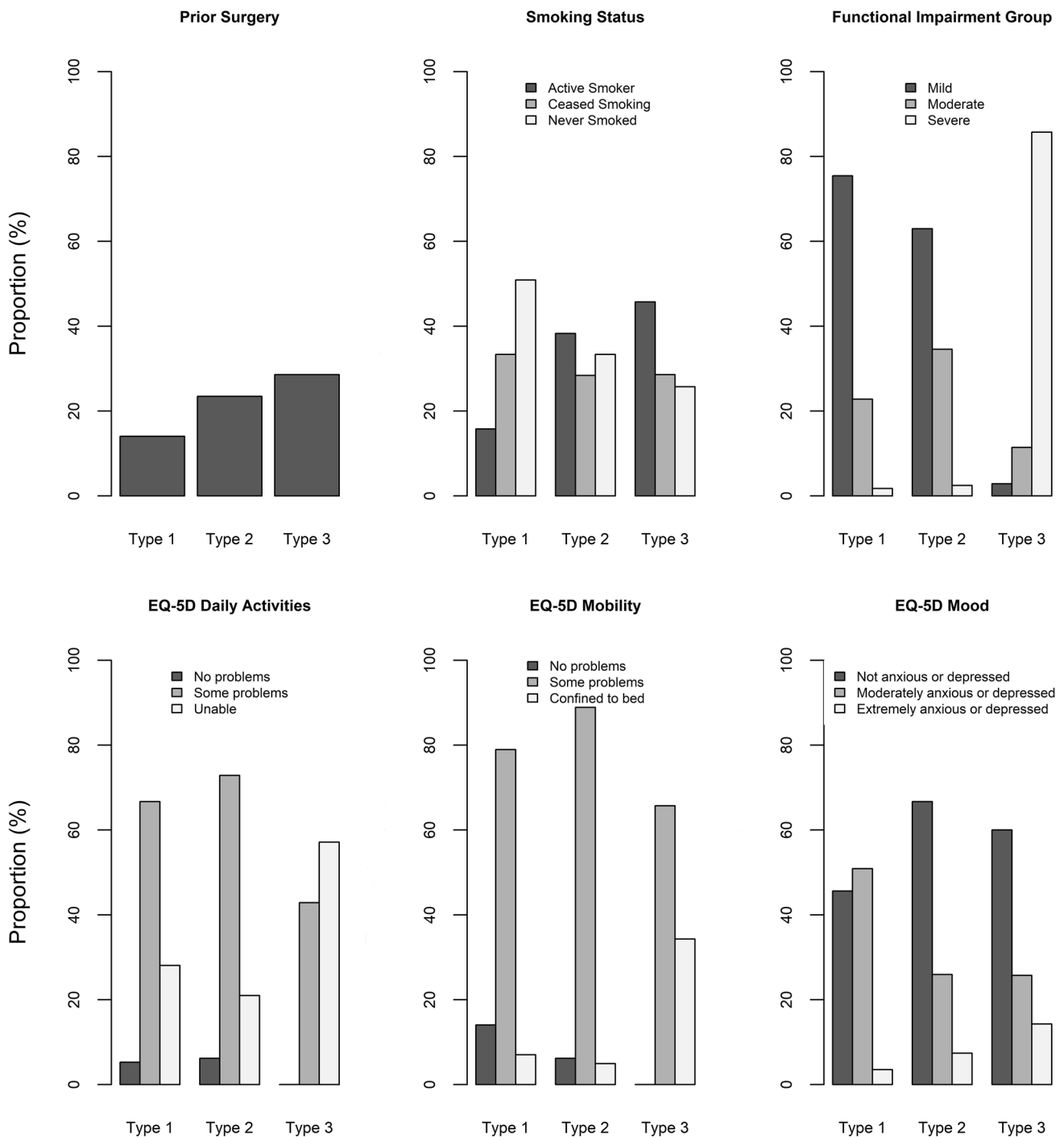
**Fig. 3** Scatterplots demonstrating the hallmarks of the three different clusters (Type 1–3) of objective functional impairment (OFI) in terms of continuous variables. Important hallmarks demonstrated are the markedly lower HRQOL and subjective disability for Type 3 OFI, as

well as the difference in body mass index between Types 1 and 2 OFI. *HRQOL* health-related quality of life, *EQ-5D* EuroQOL five-dimensions questionnaire

learning-based methods have the potential to suggest personalized “expected” cut-offs for each individual patient, based on sociodemographics, as has been alluded to in the initial validation of the 5R-STs in the spinal population [1, 37, 38]. Once a personalized cut-off has been established for the binary presence or absence of OFI, some form of impairment grading, that again takes into account sociodemographics, should be carried out, which the clustering algorithm developed in this study can do. Furthermore, machine learning methods in combination

with motion tracking-based 5R-STs assessment [39] could lead to more intuitive and automated integration of objective functional testing in clinical practice. In future, it may become possible to immediately calculate OFI, in contrast to other technological advances such as robotics, imaging, or neuronavigation, as algorithms can run server-side and even be applied on mobile devices, applications are far-reaching even in rural areas where patients cannot easily travel for in-person appointments [40].





**Fig. 4** Boxplots demonstrating the hallmarks of the three different clusters (Types 1–3) of objective functional impairment (OFI) in terms of categorical variables. Important hallmarks demonstrated are the steadily increasing rate of prior surgery, active smoking, func-

tional impairment, and extreme anxiety and depression symptoms when comparing the three clusters. *EQ-5D* EuroQOL five-dimensions questionnaire

## Limitations

Although we used prospectively collected data exclusively, this presents only single-centre data. Therefore, generalizability of our findings and specifically of the identified

clusters of OFI require further external validation before making the model available (e.g. in a web-app) and applying it in clinical practice elsewhere. However, the data that were used (preoperative sociodemographic parameters and 5R-STS testing) are not centre-specific (such as e.g. surgical

**Table 3** Comparative analysis of the three types of objective functional impairment identified in the clustering analysis by means of the variables that was not considered within the clustering analysis (Unclustered Parameters)

Parameter	Type 1 impairment	Type 2 impairment	Type 3 impairment	<i>P</i>
<i>N, (%)</i>	57 (32.9)	81 (46.8)	35 (20.2)	
<i>Unclustered parameters</i>				
Prior surgery, <i>n (%)</i>	8 (14.0)	19 (23.5)	10 (28.6)	0.211
<i>Indication for surgery, n (%)</i>				
Lumbar disc herniation	42 (73.7)	58 (71.6)	27 (77.1)	
Stenosis	10 (17.5)	17 (21.0)	2 (5.7)	
Spondylolisthesis	3 (5.3)	3 (3.7)	1 (2.9)	
Chronic low back pain	2 (3.5)	3 (3.7)	5 (14.3)	
<i>History of complaints, n (%)</i>				
< 6 wks	2 (3.5)	4 (4.9)	1 (2.9)	0.714
6 wks.–3 months	9 (15.8)	13 (16.0)	7 (20.0)	
6 months–1 year	21 (36.8)	19 (23.5)	9 (25.7)	
> 1 year	25 (43.9)	45 (55.6)	18 (51.4)	
<i>Index level, n (%)</i>				
L2–L3	1 (1.8)	4 (4.9)	1 (2.9)	0.964
L3–L4	3 (5.3)	5 (6.2)	2 (5.7)	
L4–L5	24 (42.1)	36 (44.4)	15 (42.9)	
L5–S1	29 (50.9)	36 (44.4)	17 (48.6)	
<i>Highest level of education, n (%)</i>				
Elementary school	0 (0.0)	3 (3.7)	0 (0.0)	0.649
High school	29 (50.9)	36 (44.4)	15 (42.9)	
Higher education	26 (45.6)	39 (48.1)	19 (54.3)	
(Post-)doctoral	2 (3.5)	3 (3.7)	1 (2.9)	
<i>Type of work, n (%)</i>				
Student	1 (1.8)	0 (0.0)	0 (0.0)	0.179
Houseworker	5 (8.8)	2 (2.5)	0 (0.0)	
Employed	34 (59.6)	47 (58.0)	25 (71.4)	
Self-employed	7 (12.3)	16 (19.8)	5 (14.3)	
On benefits	1 (1.8)	3 (3.7)	3 (8.6)	
Retired	7 (12.3)	6 (7.4)	1 (2.9)	
Jobless	2 (3.5)	7 (8.6)	1 (2.9)	
<i>Analgesic medication, n (%)</i>				
Not regularly	9 (15.8)	9 (11.1)	3 (8.6)	0.452
At least weekly	3 (5.3)	11 (13.6)	3 (8.6)	
Daily	45 (78.9)	61 (75.3)	29 (82.9)	
<i>Satisfied with current symptoms, n (%)</i>				
Yes	1 (1.8)	0 (0.0)	1 (2.9)	0.248
Neutral	0 (0.0)	3 (3.7)	0 (0.0)	
No	56 (98.2)	78 (96.3)	34 (97.1)	
<i>Ability to work, n (%)</i>				
Fully able	17 (29.8)	20 (24.7)	5 (14.3)	0.467
Limited	8 (14.0)	12 (14.8)	8 (22.9)	
Unable	32 (56.1)	49 (60.5)	22 (62.9)	
<i>EQ-5D mobility, n (%)</i>				
No problems	8 (14.0)	5 (6.2)	0 (0.0)	<0.001*
Some problems	45 (78.9)	72 (88.9)	23 (65.7)	
Confined to bed	4 (7.0)	4 (4.9)	12 (34.3)	
<i>EQ-5D selfcare, n (%)</i>				
No problems	29 (50.9)	44 (54.3)	10 (28.6)	0.010*

**Table 3** (continued)

Parameter	Type 1 impairment	Type 2 impairment	Type 3 impairment	<i>P</i>
Some problems	28 (49.1)	37 (45.7)	23 (65.7)	
Unable	0 (0.0)	0 (0.0)	2 (5.7)	
<i>EQ-5D daily activities, n (%)</i>				0.003*
No problems	3 (5.3)	5 (6.2)	0 (0.0)	
Some problems	38 (66.7)	59 (72.8)	15 (42.9)	
Unable	16 (28.1)	17 (21.0)	20 (57.1)	
<i>EQ-5D pain, n (%)</i>				0.003*
No pain or discomfort	2 (3.5)	3 (3.7)	0 (0.0)	
Moderate pain or discomfort	22 (38.6)	30 (37.0)	2 (5.7)	
Extreme pain or discomfort	33 (57.9)	48 (59.3)	33 (94.3)	
<i>EQ-5D mood, n (%)</i>				0.012*
Not anxious or depressed	26 (45.6)	54 (66.7)	21 (60.0)	
Moderately anxious or depressed	29 (50.9)	21 (25.9)	9 (25.7)	
Extremely anxious or depressed	2 (3.5)	6 (7.4)	5 (14.3)	
EQ-5D index, mean (SD)	0.35 (0.27)	0.40 (0.30)	0.13 (0.23)	<0.001*
EQ-5D thermometer, mean (SD)	43.89 (16.79)	50.77 (16.77)	42.80 (18.91)	0.022*
NRS back pain, mean (SD)	6.54 (2.67)	6.36 (2.16)	7.11 (2.21)	0.284
NRS leg pain, mean (SD)	7.86 (1.41)	7.35 (1.82)	7.63 (2.34)	0.261
Oswestry disability index, mean (SD)	46.98 (15.90)	45.70 (15.81)	58.57 (15.27)	<0.001*
Roland-Morris disability questionnaire, mean (SD)	13.07 (5.49)	12.89 (4.86)	16.46 (3.74)	0.001*

*SD* standard deviation, *EQ-5D* EuroQOL five-dimensions questionnaire, *NRS* Numeric rating scale

\* $p \leq 0.05$

treatment or length of stay), and the 5R-STs has been established as having extremely high inter-rater reliability [1, 23]. Inclusion of further parameters from patient history and clinical examination could possibly increase the distinctness of the clusters even further. However, this would come at the cost of clinical usability and parsimony of the algorithm and

**Table 4** Qualitative overview of the hallmarks of the three types of impairment that was identified through unsupervised analysis

Domain	Type 1 impairment	Type 2 impairment	Type 3 impairment
5R-STs test time	↓	↓	↑
Body mass index	↓	↑	↔
Gender	♀	♀	♂
Smoking	↓	↔	↑
Subjective functional impairment	↔	↔	↑
Depression & anxiety	↔	↑	↑↑
Mobility, ADL	↔	↔	↓↓
Age	↔	↔	↔
Pain	↔	↔	↔
History of complaints	↔	↔	↔
Work status & type	↔	↔	↔

5R-STs five-repetition sit-to-stand test, ADL activities of daily life

derived classification of OFI. Currently, only variables that are easily and objectively assessable such as age, BMI, and gender are included in the model, which enables clinical application in under one minute. Although we included a comparatively large and homogenous cohort of patients with OFI, a larger number of patients would likely also lead to an increase in generalizability and distinctness of the clusters. Lastly, the model was not tested in separate diagnoses such as chronic low back pain and spondylolisthesis due to lacking statistical power for such subgroup analyses. However, the classification of OFI based on our model is independent of diagnosis (i.e. it is not a factor considered in this cluster analysis), and in addition, our analysis of unclustered parameters demonstrated that there is no interaction between diagnosis and cluster assignment, indicating robustness against different diagnostic categories.

## Conclusions

In this study, we demonstrate that unsupervised machine learning techniques, in combination with the 5R-STs, identified three distinct clusters of patients with OFI that may represent a more holistic and objective clinical classification of patients than test times and baseline severity stratifications

alone by taking into account individual patient characteristics. These findings may in future be integrated with higher levels of automation into clinical practice and may then also have diagnostic, prognostic, and predictive implications for surgical and nonsurgical treatment of degenerative spinal conditions.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00586-023-08070-z>.

**Acknowledgements** The authors are grateful to all participating patients and to Femke Beusekamp, BSc and Nathalie Schouman for study coordination and data collection. We also thank Marlies P. de Wispelaere, MSc for her efforts in clinical informatics.

**Author contributions** VES and VS analyzed the data, drafted and revised the manuscript. AMK, VS, WPV, and MLS revised the manuscript critically for important intellectual content. VES, AMK, VS, WPV, and MLS made substantial contributions to the conception and design of the study and provided final approval of the version to be published. VES, AMK, VS, and MLS were involved in the acquisition of data. All authors read and approved the final manuscript.

**Funding** Open access funding provided by University of Zurich. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declarations

**Conflict of interest** The authors declare that the article and its content were composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Staartjes VE, Schröder ML (2018) The five-repetition sit-to-stand test: evaluation of a simple and objective tool for the assessment of degenerative pathologies of the lumbar spine. *J Neurosurg Spine* 29:380–387. <https://doi.org/10.3171/2018.2.SPINE171416>
2. Gautschi OP, Smoll NR, Corniola MV et al (2016) Validity and reliability of a measurement of objective functional impairment in lumbar degenerative disc disease: the timed up and go (TUG) test. *Neurosurgery* 79:270–278. <https://doi.org/10.1227/NEU.0000000000001195>
3. Gautschi OP, Corniola MV, Schaller K et al (2014) The need for an objective outcome measurement in spine surgery—the timed-up-and-go test. *Spine J* 14:2521–2522. <https://doi.org/10.1016/j.spinee.2014.05.004>
4. Stienen MN, Ho AL, Staartjes VE et al (2019) Objective measures of functional impairment for degenerative diseases of the lumbar spine: a systematic review of the literature. *Spine J* 19:1276–1293. <https://doi.org/10.1016/j.spinee.2019.02.014>
5. Jones SE, Kon SSC, Canavan JL et al (2013) The five-repetition sit-to-stand test as a functional outcome measure in COPD. *Thorax* 68:1015–1020. <https://doi.org/10.1136/thoraxjnl-2013-203576>
6. Crook S, Büsching G, Schultz K et al (2017) A multicentre validation of the 1-min sit-to-stand test in patients with COPD. *Eur Respir J* 49:1601871. <https://doi.org/10.1183/13993003.01871-2016>
7. Gautschi OP, Joswig H, Corniola MV et al (2016) Pre- and postoperative correlation of patient-reported outcome measures with standardized timed Up and Go (TUG) test results in lumbar degenerative disc disease. *Acta Neurochir (Wien)* 158:1875–1881. <https://doi.org/10.1007/s00701-016-2899-9>
8. Guyatt GH, Sullivan MJ, Thompson PJ et al (1985) The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J* 132:919–923
9. Stienen MN, Smoll NR, Joswig H et al (2017) Influence of the mental health status on a new measure of objective functional impairment in lumbar degenerative disc disease. *Spine J* 17:807–813. <https://doi.org/10.1016/j.spinee.2016.12.004>
10. Joswig H, Stienen MN, Smoll NR et al (2017) Patients' preference of the timed up and go test or patient-reported outcome measures before and after surgery for lumbar degenerative disk disease. *World Neurosurg* 99:26–30. <https://doi.org/10.1016/j.wneu.2016.11.039>
11. Sosnova M, Zeitlberger AM, Ziga M et al (2020) Patients undergoing surgery for lumbar degenerative spinal disorders favor smartphone-based objective self-assessment over paper-based patient-reported outcome measures. *Spine J Off J North Am Spine Soc.* <https://doi.org/10.1016/j.spinee.2020.11.013>
12. Jakobsson M, Gutke A, Morkink LB et al (2019) Level of evidence for reliability, validity, and responsiveness of physical capacity tasks designed to assess functioning in patients with low back pain: a systematic review using the COSMIN standards. *Phys Ther* 99:457–477. <https://doi.org/10.1093/ptj/pzy159>
13. Stienen MN, Maldaner N, Joswig H et al (2019) Objective functional assessment using the “timed up and go” test in patients with lumbar spinal stenosis. *Neurosurg Focus* 46:E4. <https://doi.org/10.3171/2019.2.FOCUS18618>
14. Tosic L, Goldberger E, Maldaner N et al (2020) Normative data of a smartphone app-based 6-minute walking test, test-retest reliability, and content validity with patient-reported outcome measures. *J Neurosurg Spine* 33(4):480–489. <https://doi.org/10.3171/2020.3.SPINE2084>
15. Klukowska AM, Schröder ML, Stienen MN, Staartjes VE (2020) Objective functional impairment in lumbar degenerative disease: concurrent validity of the baseline severity stratification for the five-repetition sit-to-stand test. *J Neurosurg Spine* 33:4–11. <https://doi.org/10.3171/2019.12.SPINE191124>
16. Stienen MN, Smoll NR, Joswig H et al (2017) Validation of the baseline severity stratification of objective functional impairment in lumbar degenerative disc disease. *J Neurosurg Spine* 26:598–604. <https://doi.org/10.3171/2016.11.SPINE16683>
17. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25:44–56. <https://doi.org/10.1038/s41591-018-0300-7>
18. Snyderman R (2012) Personalized health care: from theory to practice. *Biotechnol J* 7:973–979. <https://doi.org/10.1002/biot.201100297>
19. Alashwal H, El Halaby M, Crouse JJ et al (2019) The application of unsupervised clustering methods to alzheimer's disease. *Front Comput Neurosci* 13:31. <https://doi.org/10.3389/fncom.2019.00031>

20. Miller DJ, Wang Y, Kesidis G (2008) Emergent unsupervised clustering paradigms with potential application to bioinformatics. *Front Biosci J Virtual Libr* 13:677–690. <https://doi.org/10.2741/2711>
21. Ficiarà E, Boschi S, Ansari S et al (2021) Machine learning profiling of alzheimer's disease patients based on current cerebrospinal fluid markers and iron content in biofluids. *Front Aging Neurosci* 13:607858. <https://doi.org/10.3389/fnagi.2021.607858>
22. Ames CP, Smith JS, Pellisé F et al (2019) artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine* 44:915–926. <https://doi.org/10.1097/BRS.0000000000002974>
23. Staartjes VE, Beusekamp F, Schröder ML (2019) Can objective functional impairment in lumbar degenerative disease be reliably assessed at home using the five-repetition sit-to-stand test? A prospective study. *Eur Spine J* 28:665–673. <https://doi.org/10.1007/s00586-019-05897-3>
24. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66:271–273
25. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8:141–144
26. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 33:337–343
27. Peasgood T, Brazier J, Papaioannou D (2012) A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety. *HEDS Discuss Pap* 1215
28. Kowarik A, Templ M (2016) Imputation with the R Package VIM. *J Stat Softw* 74(7):1–16. <https://doi.org/10.18637/jss.v074.i07>
29. R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
30. Hartigan JA, Wong MA (1979) A k-means clustering algorithm. *JSTOR Appl Stat* 28:100–108
31. Siccoli A, Staartjes VE, de Wispelaere MP, Schröder ML (2018) Gender differences in degenerative spine surgery: Do female patients really fare worse? *Eur Spine J*. <https://doi.org/10.1007/s00586-018-5737-3>
32. Kim H-J, Suh B-G, Lee D-B et al (2013) Gender difference of symptom severity in lumbar spinal stenosis: role of pain sensitivity. *Pain Physician* 16:E715–723
33. Racine M, Tousignant-lafamme Y, Kloda LA et al (2012) A systematic literature review of 10 years of research on sex/gender and pain perception—Part 2: Do biopsychosocial factors alter pain sensitivity differently in women and men? *Pain* 153:619–635. <https://doi.org/10.1016/j.pain.2011.11.026>
34. Pochon L, Kleinstück FS, Porchet F, Mannion AF (2016) Influence of gender on patient-oriented outcomes in spine surgery. *Eur Spine J* 25:235–246. <https://doi.org/10.1007/s00586-015-4062-3>
35. Joswig H, Stienen MN, Smoll NR et al (2017) Effects of smoking on subjective and objective measures of pain intensity, functional impairment, and health-related quality of life in lumbar degenerative disk disease. *World Neurosurg* 99:6–13. <https://doi.org/10.1016/j.wneu.2016.11.060>
36. Brathwaite R, Smeeth L, Addo J et al (2017) Ethnic differences in current smoking and former smoking in the Netherlands and the contribution of socioeconomic factors: a cross-sectional analysis of the HELIUS study. *BMJ Open* 7:e016041. <https://doi.org/10.1136/bmjopen-2017-016041>
37. Obermeyer Z, Emanuel EJ (2016) Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 375:1216–1219. <https://doi.org/10.1056/NEJMp1606181>
38. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. *N Engl J Med* 380:1347–1358. <https://doi.org/10.1056/NEJMr1814259>
39. Ejupi A, Brodie M, Gschwind YJ et al (2015) Kinect-based five-times-sit-to-stand test for clinical and in-home assessment of fall risk in older people. *Gerontology* 62:118–124. <https://doi.org/10.1159/000381804>
40. Munakomi S (2019) Letter to the Editor. Reappraising role of clinical evaluations in degenerative lumbar spine pathologies. *J Neurosurg Spine* 30:860–861. <https://doi.org/10.3171/2018.10.SPINE181282>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.