On clustering patients with low back pain

Hanneke van der Hoef

Abstract The current study contributes to the search of identifying subgroups of patients with low back pain by using *clara*: clustering large applications. Different from prior studies, a dimension reduction is provided by selecting key variables found in the literature. In addition, external instead of internal validation criteria are followed. Five groups are identified, which are characterized as: (1) pain has spread down into the legs (2) acute, intense low back pain which is likely to be aggravated by work (3) acute intense low back pain, not aggravated by work, and sleeping problems (4) no (activity) limitations, good recovery rate (5) chronic (i.e. more than 3 months) low back pain with a bad prognosis. Limitations and recommendations are discussed.

1 Introduction

To better understand the mechanisms underlying the heterogeneity of patients with low back pain, and consequently to be able to provide tailored treatment, it is of interest to identify subgroups of patients with low back pain. Van Mechelen

Hanneke van der Hoef
University of Groningen, The Netherlands
✉ h.van.der.hoef@rug.nl

Archives of Data Science, Series B
(Online First)
KIT Scientific Publishing
Vol. 1, No. 1, 2019

DOI 10.5445/KSP/1000085952/02
ISSN 2510-0564

and Vach (2018) have provided an open data set of lower back pain patients which is analyzed in this contribution. In past years, several studies tried to identify subgroups using cluster analysis (e.g. Hirsch et al (2014); Nielsen et al (2017); Rabey et al (2015)). However, these studies vary greatly in their procedures, as well as their outcomes. It is a hard task to assess which clustering provides the best results, because there is no 'true' sub-grouping (i.e. there are only reasons to assume that a sub-grouping is present, but the groups are not definite). In addition, various different internal and external validity indices can be used to assess the quality of a clustering.

The current study contributes to the search of identifying groups of patients with low back pain by using cluster analysis. First, a dimension reduction is provided by selecting key variables that show -according to prior studies- high variation among participants with low back pain. Next, rather than relying solely upon internal validity indices, a clear interpretation and differentiation among the clusters on symptoms and longitudinal outcome measures are used as criteria to select the final clustering. One of the underlying ideas here is that clinical interpretability should be kept in mind at all times.

2 Clustering procedure

Cluster analysis consists of many steps, each requiring the user to make decisions that will influence the results. Major steps in cluster analysis are: variable selection, the selection of a clustering method, determining the number of clusters, and cluster validation (Walesiak and Dudek, 2010). All steps are considered in an iterative process for the current study, and discussed below.

2.1 Variable selection

One of the main challenges that the user of cluster analysis is faced with is high-dimensionality of the data. Often, as in this case, one has access to many variables, and desires to cluster the objects or participants on these variables. But, when clustering a large amount of variables, the analysis is sensitive to the 'curse of dimensionality' (Bellman, 1961), which refers to the phenomenon that

a clustering technique, which works well in low-dimensionality, often performs poorly when the dimensionality increases (Steinbach et al, 2004).

A main reason why high-dimensionality is problematic in cluster analysis is caused by the distance metric. Most clustering methods use distances as the (dis)similarity criterion to form the clusters and thus depend heavily on the distance metric. However, many distances behave 'surprisingly' in high-dimensional data, and the concept of distance may even lose its qualitative meaning (Aggarwal et al, 2001). In short, this is due to 'distance concentration', i.e. the tendency in high-dimensional data that the points are all far away from the centre with little difference remaining between the points (Radavanović et al, 2015). To overcome the problem of distance concentration, a possible solution is to reduce the dimensionality.

Several methods have been proposed to lower the dimensionality, such as dimension reduction by statistical techniques (e.g. principal component analysis, singular-value decomposition and multiple correspondence analysis) or to remove variables that are not of interest, highly correlated or show little variation among objects. However, the statistical techniques that can be used to reduce the dimensionality rely upon the assumption that all objects lie within the same, full space. Yet, it is often, as in this case, more appropriate to assume that objects lie within different subspaces. When the same space assumption is not met, cluster analysis after dimension reduction by statistical techniques is inappropriate (Steinbach et al, 2004).

For the current study then, a dimension reduction is obtained by selecting key variables based on indications from prior studies. More specifically, the Nielsen et al (2016) paper was used to identify variables that showed evident differences among the patient groups. These variables were explicitly mentioned as characteristics that discriminate among groups and/or showed clear differences among groups in the plots. Next, additional variables were used that differed significantly among groups in several related papers (Eirikstoft and Kongsted, 2014; Kongsted et al, 2014; Myers et al, 2008). This led to the selection of 13 key variables, which are summarized in Table 1. All items are qualitative; either dichotomous (10) or ordinal (3).

In accordance with Nielsen et al (2016) only single-items were used for the clustering. Single-items provide a more detailed view than summary statistics. Additionally, this procedure matches the aim of finding key variables, which can be used in clinical practice to group patients without having to collect many items.

Table 1: Description of the 13 selected key variables.

Item	Description	Scale	Retrieved from
rm150	sleep less well	dichotomous	Nielsen et al (2016)
rm160	avoid heavy jobs around the house	dichotomous	Nielsen et al (2016)
rm170	more irritable and tempered than usual	dichotomous	Nielsen et al (2016)
mdi8	restless, subdued or slowed down	ordinal (3)	Nielsen et al (2016)
tlda0	lbp frequency: > 30 days in last year	dichotomous	Nielsen et al (2016)
domin_lbp	lbp is dominating	dichotomous	Nielsen et al (2016)
start10	pain has spread down into legs	dichotomous	Nielsen et al (2016)
start30	only walk short distances due to lbp	dichotomous	Nielsen et al (2016)
start40	dressed more slowly last 2 weeks	dichotomous	Nielsen et al (2016)
start90	bothered by pain last 2 weeks	dichotomous	Nielsen et al (2016)
fabq70	work aggravated pain	ordinal (3)	Eirikstoft and Kongsted (2014)
dlva0	duration of current lbp	ordinal (4)	Eirikstoft and Kongsted (2014)
okom0	(negative) recovery belief	dichotomous	Myers et al (2008)

Next, 3 participants were excluded because they had missing values on all 13 variables (participants with ID: 74, 328, and 674). Consequently, $n = 925$ participants and $p = 13$ variables remained for the cluster analysis.

2.2 Selecting a method

A clustering method is required that is suitable for dichotomous and ordinal variables. In addition, the method should be able to handle missing data, and preferably be robust, i.e. be as little as possible influenced by outliers and noise. K-medoids by *clara* (Clustering LARge Applications) in the R (R Core Team, 2013) package *cluster* (Maechler et al, 2017) satisfies these requirements, and is therefore chosen as the clustering method. The function *clara* is an extended version of *pam* (partitioning around medoids). It is a sample-based algorithm, which creates sub-datasets and hence requires less computational time and storage (Kaufman and Rousseeuw, 1990). Additionally, different from for instance k-means, the k-medoids algorithm uses real observations as the cluster centers. These are subsequently useful in aiding the interpretation of the clusters. Moreover, because *clara* uses representative objects (medoids) instead of centroids, it is less sensitive to outliers compared to k-means clustering, and hence considered more robust (Park and Jun, 2009).

2.3 Determining the number of clusters and cluster validation

Choosing the number of clusters should not be based solely on data without user input (Hennig, 2015). Hence, to assess the optimal number of clusters, validation of the clusterings is based on both internal and external properties.

Internal validity

Many indices have been proposed to assess the internal validation of clusterings. However, no consensus has been reached as to which index reflects the quality of the clustering best. This is -at least partly- caused by the fact that a ‘good’ clustering depends upon the specific aim of the cluster analysis. For the current study the aim is to find clusters that differentiate on their symptoms and prognosis.

Overlapping clusters are allowed, but a reasonable similarity should be present within clusters. To assess the internal quality, three measures are selected that reflect the connectedness, compactness and separation of the clusters (Brock et al, 2008).

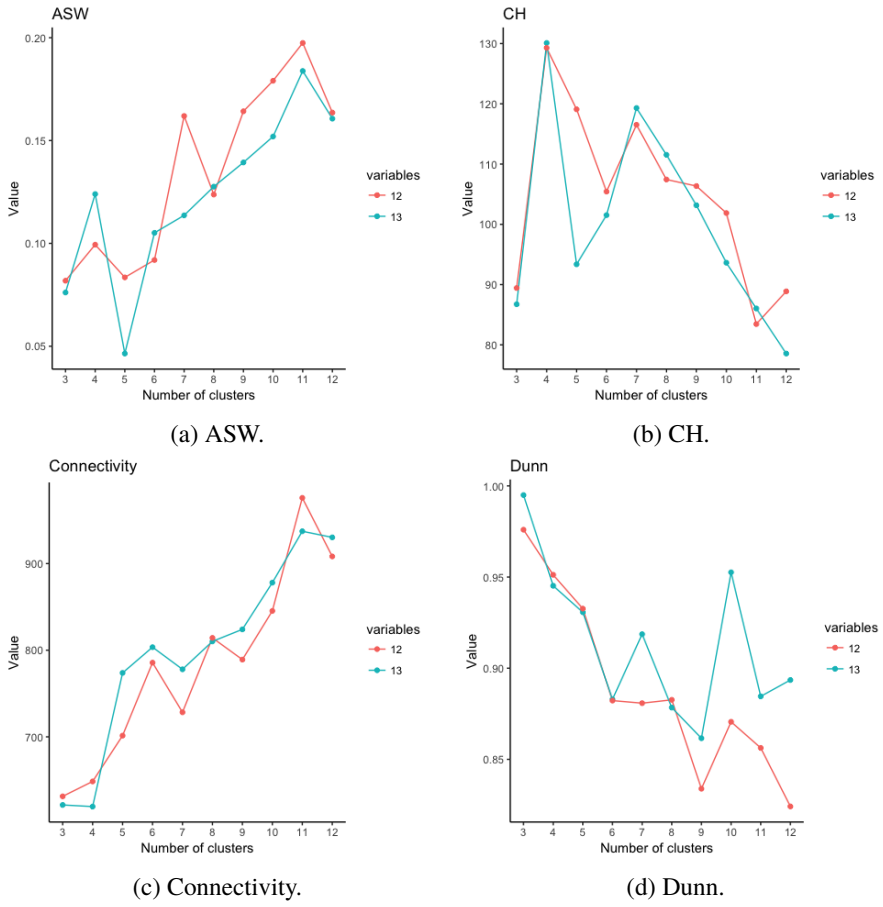
Connectedness is measured by *connectivity*, which measures the cluster homogeneity, i.e. the extent to which observations that are j^{th} nearest neighbours in the data space, are placed together in a cluster (Brock et al, 2008). Connectivity ranges from [0, ∞], in which lower values are favourable.

Compactness quantifies the degree of within-cluster variance, whereas separation quantifies the degree of between-cluster variance, usually by measuring the distance between cluster centroids, or -here- medoids (Brock et al, 2008). Compactness and separation reflect opposite trends, i.e. when the number of clusters increases, compactness likely increases and separation decreases (Brock et al, 2008). Two commonly used indices that combine compactness and separation into one measure are the average silhouette width (ASW: Rousseeuw, 1987) and the Dunn index (Dunn, 1974). ASW measures the "degree of confidence" of the cluster assignment and is bounded by [-1, 1] in which higher values indicate more confidence and thus are favourable (Brock et al, 2008). The Dunn index is the ratio between the minimum distance between observations in different clusters, and the maximum distance between observations in the same cluster (Brock et al, 2008). The Dunn index is bounded by [0, ∞], in which higher values are favourable.

In addition, the Calinski and Harabasz (CH) index (Calinski and Harabasz, 1974) is computed to assess the optimal number of clusters according to the ratio of between and within sum-of-squares of k clusters. The CH index showed one of the best results in searching the optimal number of clusters in the well-known comparative study by Milligan and Cooper (1985), making it a popular index to assess the number of clusters. The CH index should be maximized.

Figure 1 shows the values for the ASW, CH, Connectivity and Dunn index for $k = 3$ to $k = 12$ clusters, as retrieved by clustering with *clara* on the 13 selected key variables (blue lines). The optimal values are the maximum values for ASW, CH and Dunn, and the minimum value for Connectivity. Accordingly, the optimal number of clusters is $k = 11$ for the ASW, $k = 4$ for CH and Connectivity, and $k = 3$ for the Dunn index. Thus, there is no agreement according to these four indices as to what the optimal number of clusters is. For the time being, the

three "winning solutions", i.e. $k = 3$, $k = 4$ and $k = 11$ are chosen to validate further on the external reference.



(a) ASW.

(b) CH.

(c) Connectivity.

(d) Dunn.

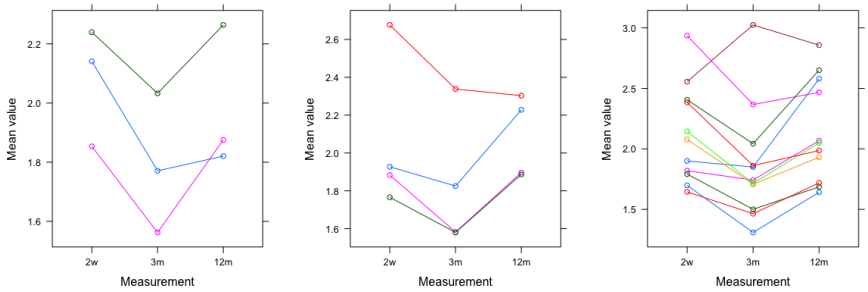
Figure 1: Optimal number of clusters according to ASW, CH, Connectivity and Dunn index for clustering on $p = 13$ variables (blue) and after removing item *rm160*, thus with $p = 12$ variables (pink). *Note*: Optimal value is maximum for ASW, CH, and Dunn index, and minimum for connectivity.

External validity

The external validity of the clusterings is evaluated on clarity and meaning of the clusters, i.e. participants within a cluster are expected to have similar

symptoms and a similar prognosis, reflected by longitudinal measures on (1) global perceived health, (2) disability and (3) pain intensity, all measured at 2 weeks, 3 months and 12 months after baseline.

The three 'winning' clusterings are further evaluated on their clarity and differentiation on the longitudinal outcome measures. An example is provided in Figure 2. In Figure 2, the mean scores on global perceived improvement at 2 weeks, 3 months and 12 months after baseline are plotted for the clusters. First of all, the 11-cluster solution is hard to asses; it is difficult to interpret the differences among the clusters. This was the case for all longitudinal measures. In addition, the 11 clusters did not show clear and significant differences on the (baseline) symptoms, making it a poor outcome in terms of the specific clustering aim. Hence, the 11-cluster solution is discarded. When comparing the remaining 2 solutions, it is found that the 3-cluster solution shows clearer differences among the groups and a clearer interpretation (in the 4-cluster solution, the difference between cluster 2 and 3 was difficult to grasp). Therefore, it was decided to continue with the three-cluster solution.



(a) 3 clusters. (b) 4 clusters. (c) 11 clusters.

Figure 2: Mean scores on Global Perceived Improvement at 2 weeks, 3 months and 12 months after baseline for three cluster solutions, with respectively k = 3, 4 and 11 clusters. *Note*: Of the n = 925 patients, n = 503 had no missing values, n = 222 had 1 missing value, n = 95 had 2 missing values, and n = 105 had all three values missing. The participants with missing values on a measurement are excluded from calculating the group-mean on that measurement.

2.4 Variable impact assessment

Next, the impact of each variable on the partitioning was assessed with the Adjusted Rand Index (ARI: Hubert and Arabie (1985)). The ARI is a commonly used index to compare clusterings. Here, it is used to compare the full clustering with the clustering after omitting one variable at a time to quantify how much influence each variable has on the clustering. A value of one indicates identical partitions, and thus no influence of the variable on the clustering. A value of zero indicates independent partitions, and thus an extremely high variable impact. While values within this range may be more difficult to pinpoint, ARI > .75 is considered notably high, and worthwhile to consider exclusion of the variable. Table 2 presents ARI-values after omitting one variable at a time. As can be seen in the table, one variable clearly stands out in terms of a high value, and thus a low variable impact: *rm160* (ARI = .86). Looking at the contingency table, it is found that 54 of the 925 participants (6%) are allocated to a different cluster when omitting this variable. Since a difference in allocation of six per cent of the participants is considered a minor change, this variable was excluded. The ARI of all remaining variables is well below .75, with *tlda*0 now having the lowest variable impact (ARI = .55).

Table 2: Variable impact as determined by the Adjusted Rand Index (ARI) for the full partition and the partition after omitting that variable.

Item	k = 3 (full model)	k = 3 (after exclusion of *rm160*)
rm150	.43	.45
rm160	.86	n.a.
rm170	.41	.38
mdi8	.12	.23
tlda0	.48	.55
domin_lbp	.47	.48
start10	.23	.10
start30	.39	.13
start40	.25	.12
start90	.28	.25
fabq70	.34	.38
dlva0	.10	.14
okom0	.46	.48

2.5 Revision of choosing the optimal number of clusters

After excluding variable $rm160$, the optimal number of clusters is reconsidered by both internal and external validity criteria. The pink lines in Figure 2 for the clustering based on the remaining 12 variables, show quite similar patterns compared to the clustering on the 'full' set of 13 variables, although some differences are present. As to the optimal number of clusters, results are equal for ASW, with $k = 11$, CH with $k = 4$ and the Dunn index with $k = 3$. A small change occurred for Connectivity; the preference shifted from $k = 4$ to $k = 3$.

Because the internal validity indices do not provide an unambiguous 'winner' and there is no theoretical motivation to rely upon one index, it is decided to choose the final clustering on the clearest interpretation and differentiation on the longitudinal measures. Briefly, for a large number of clusters it is difficult to distinguish patterns among the groups. Vice versa, a low number of clusters may cause an unnecessary loss of information when groups are merged together that in fact do show different patterns on their symptoms and longitudinal measures. A balance between these two considerations is found for a clustering with 5 groups. The groups in the 5-cluster solution still showed clear differences among groups on their prognosis and symptoms, while the two 'extra' groups compared to the minimal number of clusters ($k = 3$) provided two meaningful, well-distinguishable additional groups. In addition, although 5 is not the optimal number of clusters according to any of the four indices discussed before, the 5-cluster solution does show relatively good values on all four measures: ASW = 0.07, CH = 119, Connectivity = 701.40, and Dunn index = 0.93. Therefore, the final clustering is retrieved from $clara$ with $p = 12$ variables. This resulted in $k = 5$ groups, which are interpreted in the next section.

3 Interpretation of the clustering results

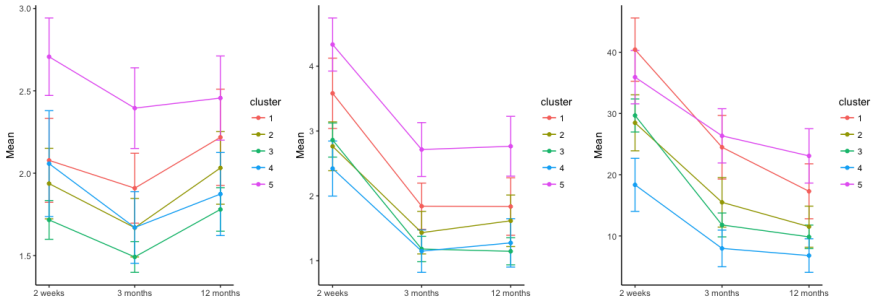
Table 3 presents descriptive characteristics for the five clusters. In addition, Figure 3 shows the mean longitudinal outcomes for the five clusters on global perceived improvement, low back pain intensity and disability.

Table 3: Baseline characteristics of the five clusters (1/2).

Descriptor	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Size (*n*)	108	165	377	106	169
Sex: female	53%	39%	46%	31%	53%
Age: *mean (SD)*	46 (11)	42 (11)	44 (11)	42 (11)	41 (13)
Frequency: 30+ days of lbp in past year	44%	3%	14%	15%	81%
Duration of LBP[a]					
0-2 weeks	43%	93%	82%	67%	0%
2-4 weeks	25%	7%	15%	20%	5%
1-3 months	16%	0%	3%	13%	33%
3+ months	17%	0%	0%	0%	63%
Pain intensity (0-10)[b]: *mean (SD)*	6.1 (2.0)	7.1 (2.0)	7.1 (1.7)	4.7 (2.2)	6.2 (2.0)
Self-rated general health (0-100)[c]: *mean (SD)*	65 (18)	66 (22)	68 (20)	80 (16)	64 (21)
Sleeping problems[d]	70%	36%	79%	9%	54%
More irritable than usual[e]	58%	26%	40%	7%	43%
Restless, subdued, slowed down[f]					
At no time	18%	13%	28%	79%	28%
Some of the time	62%	61%	40%	21%	43%
Slightly - all the time	21%	26%	32%	0%	29%
Pain spread down into legs[g]	91%	47%	33%	18%	44%
Only walk short distances[h]	68%	46%	32%	5%	17%
Dresses more slowly last 2 weeks[g]	89%	64%	77%	21%	46%
Bothered by LBP last 2 weeks[i]	52%	86%	82%	10%	65%
Work aggravated pain[i]					
Disagree	41%	12%	71%	83%	26%
Unsure	17%	33%	17%	11%	23%
Agree	42%	55%	12%	7%	51%
Negative recovery belief[j]	16%	15%	22%	8%	74%

Note. Missing values are excluded. [a]Missing: n = 15 (2%); [b]Missing: n = 22 (2%); [c]Missing: n = 30 (3%);[d]Missing: n = 19 (2%); [e]Missing: n = 14 (2%); [f]Missing: n = 3 (0%); [g]Missing: n = 13 (1%); [h]Missing: n = 26 (3%); [i]Missing: n = 9 (1%); [i]Missing: n = 29 (3%); [j]Missing: n = 10 (1%).

Group 1 is characterized primarily by the fact that the pain has spread down into the legs (91%). Accordingly, patients in group 1 often report activity limitations, such as only being able to walk short distances (68%) and problems dressing (89%). With regard to their prognosis, patients in this group reported the highest scores on the Roland-Morris score for disability 2 weeks after baseline ($\mu = 40.0$, $SD = 25.0$). But, they also show a fairly high improvement with regard to disability and pain intensity, reflected by a decrease to $\mu = 24.1$ ($SD = 24.4$) at three months after baseline, and $\mu = 17.0$ ($SD = 19.6$) one year after baseline.



(a) Global improvement. (b) Low back pain intensity. (c) Disability.

Figure 3: Line plots displaying the mean and 95%-CI for the five clusters on the longitudinal measures. *Note*. Missing values are excluded: Missing on global perceived improvement: $n_{2w} = 282$, $n_{3m} = 198$, $n_{12m} = 247$. Missing on low back pain intensity: $n_{2w} = 307$, $n_{3m} = 197$, $n_{12m} = 244$. Missing on disability: $n_{2w} = 152$, $n_{3m} = 194$, $n_{12m} = 243$..

Group 2 is characterized by short, but intense episodes of low back pain. Almost all patients report a short (i.e. 0-2 weeks) duration of the current low back pain episode (93%), but at the same time they report high pain intensity ($\mu = 7.1$, $SD = 2.0$) and most patients are very or extremely bothered by their pain in the last two weeks (86%). The majority of the patients thinks that work aggravated the pain (55%). Most patients (85%) believed to fully recover within 3 months, and indeed, they show fairly good improvement with regard to the pain intensity and disability between 2 weeks and 3 months after baseline.

Group 3 is by far the largest group with $n = 377$. Similarly to group 2, most patients report a short duration of the current low back pain episode (82%) and a high pain-intensity ($\mu = 7.1$, $SD = 1.7$) and bothersomeness of the pain (82%). They differ from the previous group by reporting a much higher rate of sleeping

problems (79%). In addition, patients in group 3 do not consider their pain to be aggravated by their work (12%). With regard to their prognosis, their pattern is similar to group 2, although the values remain slightly lower (i.e. better) than group 2 at each measurement.

Group 4 is the smallest group ($n = 106$), and formed by a majority of men (69%). This group can be characterized as the 'best' group. Most patients report short episodes of low back pain (67%), with the lowest pain intensity ($\mu = 4.7$, $SD = 2.2$) and the highest self-rated general health ($\mu = 80$, $SD = 16$). They report almost no sleeping problems, psychological problems or activity limitations. Most patients disagree that their pain is aggravated by their work (83%) and they believe in full recovery within 3 months (92%). Accordingly, they show almost no pain intensity one year after baseline ($\mu = 1.3$, $SD = .7$) nor disability ($\mu = 6.6$, $SD = 12.2$).

Group 5 is the chronic and perhaps most problematic group. Patients in group 5 report long episodes of low back pain, with the majority reporting a current low back pain episode of 3 months or more (63%). Most patients (74%) do not believe in full recovery within three months. While they do show some improvement regarding pain intensity and disability, especially between 2 weeks and 3 months after baseline, they still report high values of pain and disability one year after baseline consultation.

Lastly, while all groups show (some) decline in disability and pain intensity, many patients report no global perceived improvement (Figure 3a). For all groups, low mean values on global improvement are reported (1 = much worse; 7 = much better). This is worrying, yet in accordance with prior studies yielding low treatment effects and bad recovery rates.

4 Discussion

With the function *clara* on 12 selected key variables, 5 clusters were identified that can be characterized as:

(1) pain in the legs, high disability, and activity limitations,

(2) acute, short-term episode(s) of low back pain with high pain intensity, and pain might be aggravated by work,

(3) largest group, short and intense episode(s) of low back pain, yet pain is not aggravated by work, sleeping problems, and best prognosis with regard to pain intensity,

(4) 'best' group, almost no (activity) limitations, mostly male, low pain intensity, high self-rated general health, and good recovery rates,

(5) chronic low back pain, moderate pain intensity, negative recovery belief, some improvement within one year after baseline but high disability, and pain intensity rates remain.

With the current study it is not claimed that the optimal clustering has been found. For example, one drawback of this study concerns the high uncertainty, reflected by high standard deviations and hence broad confidence intervals on the longitudinal measures. A different, perhaps better, clustering may have been found with for instance a different distance measure (the Manhattan or Gower distance may be more appropriate for ordinal data than the Euclidean), transformation of the variables, starting with a different set of variables and so on. However, the current study shows the results for a theory based selection of key variables, and following primarily external instead of internal validity criteria. It outlines how a sequence of decisions that have to be made in cluster analysis influence the results. Yet, with conflicting output and a lack of guidelines about when to make which decision, it is a hard task to assess what is best. The development of clear, insightful guidelines may facilitate the use of cluster analysis and result in a more unified approach towards clustering.

References

Aggarwal C, Hinneburg A, Keim D (2001) On the surprising behavior of distances in high dimensional data. In: International Conference on Database Theory (ICDT) 2001, Lecture Notes in Computer Science (LNCS) 1973, Van den Bussche J, Vianu V (eds), Springer, Berlin, p. 420–434, ISBN: 978-3-540414-56-8, DOI 10.1007/3-540-44503-X_27

Bellman R (1961) Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton, New Jersey, DOI 10.1002/nav.3800080314

Brock G, Pihur V, Datta S, Datta S (2008) clValid: An R package for cluster validation. Journal of Statistical Software 25(4):1–22, DOI 10.18637/jss.v025.i04

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Communications in Statistics 3(1):1–27, Taylor & Francis, DOI 10.1080/03610927408827101

Dunn J (1974) Well separated clusters and fuzzy partitions. Journal of Cybernetics 4(1):95–104, DOI 10.1080/01969727408546059

Eirikstoft H, Kongsted A (2014) Patient characteristics in low back pain subgroups based on an existing classification system: A descriptive cohort study in chiropractic practice. Manual Therapy 19(1):65–71, DOI 10.1016/j.math.2013.07.007

Hennig C (2015) Clustering strategy and method selection. In: Handbook of Cluster Analysis, Hennig C, Meila M, Murtagh F, Rocci R (eds), 1st edn., Chapman & Hall/CRC Press, Boca Raton, Florida, p. 703–730, ISBN: 978-1-466551-88-6

Hirsch O, Strauch K, Held H, Redaelli M, Chenot JF, Leonhardt C, Keller S, Baum E, Pfingsten M, Hildebrandt J, Basler HD, Kochen M, Donner-Banzhoff N, Becker A (2014) Low back pain patient subgroups in primary care: Pain characteristics, psychological determinants, and health care utilization. The Clinical Journal of Pain 31(12):123–132, DOI 10.1097/AJP.0000000000000100

Hubert L, Arabie P (1985) Comparing partitions. Journal of Classification 2(1):193–218, DOI 10.1007/BF01908075

Kaufman L, Rousseeuw P (1990) Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, New York, DOI 10.1002/9780470316801

Kongsted A, Vach W, Axo M, Bech R, Hestbaek L (2014) Expectation of recovery from low back pain: A longitudinal cohort study investigating patient characteristics related to the expectations and the association between expectations and 3-month outcome. Spine 39(1):81–90, DOI 10.1097/BRS.0000000000000059

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2017) cluster: Cluster Analysis Basics and Extensions, r package version 2.0.6

van Mechelen I, Vach W (2018) Cluster analyses of a target data set in the IFCS cluster benchmark data repository: Introduction to the special issue. Archives of Data Science, Series B 1(1):1–12, DOI 10.5445/KSP/1000085952/01

Milligan G, Cooper M (1985) An examination of procedures for determining the number of clusters in a data set. Psychometrika 50(2):159–179, DOI 10.1007/BF02294245

Myers S, Phillips R, Davis R, Cherkin D, Legedza A, Kaptchuk T, Hrbek A, Buring J, Post D, Connelly M, Eisenberg D (2008) Patient expectations as predictors of outcome in patients with acute low back pain. Journal of General Internal Medicine 23(2):148–153, DOI 10.1007/s11606-007-0460-5

Nielsen AM, Vach W, Kent P, Hestbaek L, Kongsted A (2016) Using existing questionnaires in latent class analysis: Should we use summary scores or single items as input? A methodological study using a cohort of patients with low back pain. Clinical Epidemiology 8:73–89, DOI 10.2147/CLEP.S103330, URL http://dx.doi.org/10.2147/CLEP.S103330

Nielsen AM, Kent P, Hestbaek L, Vach W, Kongsted A (2017) Identifying subgroups of patients using latent class analysis: Should we use a single-stage or two-stage approach? A methodological study using a cohort of patients with low back pain. Clinical Epidemiology 18(57):1–17, DOI 10.1186/s12891-017-1411-x, URL http:

//dx.doi.org/10.1186/s12891-017-1411-x

Park H, Jun C (2009) A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications 36(2):3336–3341, DOI 10.1016/j.eswa.2008.01.039

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rabey M, Slater H, O'Sullivan P, Beales D, Smith A (2015) Somatosensory nociceptive characteristics differentiate subgroups in people with chronic low back pain: A cluster analysis. PAIN 156(10):1874–1884, DOI 10.1097/j.pain.0000000000000244

Radavanović M, Nanopoulos A, Ivanović M (2015) Reverse nearest neighbors in unsupervised distance-based outlier detection. IEEE Transactions on Knowledge and Data Engineering 27(5):1369–1382, DOI 10.1109/TKDE.2014.2365790

Steinbach M, Ertöz L, Kumar V (2004) The challenges of clustering high-dimensional data. In: New Directions in Statistical Physics, Wille L (ed), Springer, Berlin, p. 273–309, ISBN: 978-3-642077-39-5, DOI 10.1007/978-3-662-08968-2_16

Walesiak M, Dudek A (2010) Finding groups in ordinal data - An examination of some clustering procedures. In: Classification as a Tool for Research., Locarek-Junge H, Weihs C (eds), Springer, Berlin, p. 185–192, ISBN: 978-3-642107-45-0, DOI 10.1007/978-3-642-10745-0_19