CHAPTER 6

Trajectories of physical functioning and their prognostic indicators; a prospective cohort study in older adults with joint pain and comorbidity

Lotte AH Hermsen
Martin Smalbrugge
Johannes C van der Wouden
Stephanie S Leone
Joost Dekker
Henriëtte E van der Horst

Published in: Maturitas 2014;78:316-22
ABSTRACT

Objectives: this study aimed to identify and characterize homogeneous subgroups of individuals with distinct trajectories of physical functioning (PF) and to examine prognostic indicators of deterioration in PF in a highly heterogeneous population of older adults with joint pain and comorbidity.

Study design: a prospective cohort study among 407 older adults with joint pain and comorbidity provided data over a period of 18 months, with 6 month time-intervals. We used latent class growth modelling (LCGM) to identify underlying subgroups (clusters) with distinct trajectories of PF. Next, we characterized these subgroups and applied multivariable logistic regression analysis to identify prognostic indicators for deterioration in PF.

Main outcome measures: we measures PF with the RAND-36 PF subscale and several potential sociodemographic, physical and psychosocial prognostic indicators.

Results: LCGM identified three clusters. Cluster 1 ‘good PF’ contained 140 participants with good baseline PF and small improvements over time. Cluster 2 ‘moderate PF’ contained 130 participants with moderate baseline PF and deterioration over time. Cluster 3 ‘poor PF’ contained 137 participants with poor baseline PF and deterioration over time. After backward selection, the final model that could best distinguish between improved participants (cluster 1) and deteriorated participants (cluster 2-3) included the following prognostic indicators: higher age, more depressive symptoms, less perceived self-efficacy and more activity avoidance.

Conclusions: older adults with joint pain and comorbidity either improved or deteriorated in PF over time. The prognostic model facilitates the classification of patients, the provision of more accurate information about prognosis and helps to narrow the focus to the high risk group of poor PF.
INTRODUCTION

Joint pain is one of the most prevalent health problems in older adults and the leading cause of deterioration in physical functioning (PF)\(^1,2\). Up to 30% of consultations in primary care are due to musculoskeletal problems, like joint pain\(^3\). Receiving information on prognosis is an important reason for patient consultation. To provide this information, clinicians need to be aware of the different trajectories of PF over time and their prognostic indicators.

It is known from previous studies on musculoskeletal disorders that various sociodemographic, physical and psychosocial factors may influence PF in older populations. However, results on prognostic models are conflicting\(^4-6\). This is probably due to the observation that most of these models were developed for single-site musculoskeletal pain (e.g. back pain, knee pain)\(^6\), while in daily practice most pain complaints manifest in multiple joints\(^7\). Furthermore, the presence of other chronic health problems (comorbidity) besides joint pain is more often rule than exception in older populations\(^8\), but still not always included in research.

The co-existence of multiple joint pain and other chronic diseases could have additional negative effects on levels of PF and the prognosis of PF\(^9\). Also, it indicates substantial heterogeneity in older populations, which should be taken into account when performing prognostic studies. However, most previous studies have provided an estimate of average change in PF over time, assuming one single trajectory that represented all individuals in the study. This approach will conceal the variety of trajectories that may occur in older populations with complex health problems. The identification of subgroups with different trajectories and their prognostic indicators may help clinicians in the provision of more accurate and individualized information to patients with joint pain and comorbidity regarding the expected course of PF. Also, it may narrow the focus to the high risk groups and support decision making regarding management of symptoms.

Therefore, this study aimed to identify and characterize homogeneous subgroups of individuals with distinct trajectories of PF and to develop a prognostic model for deterioration in PF, in a population of older adults with joint pain and comorbidity.

METHODS

Design
A prospective cohort study was conducted among 407 participants with joint pain and comorbidity. Data were collected at baseline (questionnaire and physical tests) and at 6, 12 and 18 months follow-up (questionnaires). The Medical Ethics Committee of the VU Medical
Center Amsterdam approved the study protocol and written informed consent was obtained from all participants.

Study population
Data were collected between November 2010 and April 2013. Participants were recruited from 22 general practices (GP) in the region of Amsterdam and eligible for participation if they (i) were ≥65 years, (ii) had ≥2 chronic diseases registered in the electronic medical files of the GPs, and (iii) reported joint pain on most days in the past month in at least one of eight joint pain sites: neck, back, shoulder, elbow, hand/wrist, hip, knee or ankle/foot. Participants were excluded if they lived in a nursing home, resided outside the research area, had a life threatening illness, suffered from cognitive impairments (e.g. dementia) or had insufficient knowledge of the Dutch language. Details about the study design and recruitment process have been previously published.

Outcome
Physical functioning was measured with the RAND-36 PF subscale, which asks about difficulties in a hierarchical range of 10 activities: vigorous activities, moderate activities, lift/carry groceries, climb several flights, climb one flight, bend/kneel, walk 1 km, walk 0.5 km, walk 100 m, bath/dress. Items were scored on an ordinal 3-point scale (severe, some, no limitations), recoded, summed into scale scores and transformed to a 0-100 score, with a lower score reflecting more limitations. The RAND-36 has proven to be reliable and valid in a Dutch older population.

Potential prognostic indicators
Based on available literature, the following prognostic indicators were included in the baseline questionnaire: age, gender, educational level (primary, secondary, college/university), living situation (alone, not alone), number of joint pain sites: neck, back, shoulder, elbow, wrist/hand, hip, knee and ankle/foot; score range 1-8; higher score indicates more pain sites, pain severity: 3 items of the Chronic Pain Grade (CPG); score range 0-100; higher score indicates more pain, number of chronic diseases (2, ≥3), frailty (yes, no): positive when participants met three or more of five frailty component criteria: weight loss, weakness, slowness, exhaustion, low activity, depressive symptoms: 7 items of the 14-item Hospital Anxiety and Depression Scale (HADS); score range 0-21; higher score indicates more symptoms, self-efficacy: 6-item Arthritis Self Efficacy Scale (ASES); score range 6-60; higher score indicates more self-efficacy (thus positive), activity avoidance: 5-item resting subscale of the Pain Coping Inventory (PCI); score range 5-20; higher score indicates more activity avoidance, catastrophizing: 2-item Coping Strategy Questionnaire (CSQ); score range 0-6; higher score indicates more
catastrophizing\textsuperscript{18,20}, and social support: 12-item Social Support Scale (SSS); score range 12-60; higher score indicates less perceived social support\textsuperscript{21}.

**Statistical analyses**

Whereas conventional longitudinal analyses determine only one single trajectory, latent class growth modelling (LCGM) is able to identify more underlying trajectories (clusters) that describe developmental patterns. Each identified cluster contains its own intercept (baseline value) and slope (growth value), which in LCGM are fixed to zero, to make trajectories within-clusters homogeneous and between-clusters heterogeneous\textsuperscript{22}. To identify the optimal number of clusters, we started with a single cluster model, as comparable with normal longitudinal growth modelling. Since we had four time points, we first tested this model with a quadratic component. If the quadratic component was not significant, we fitted a linear model. However, if the quadratic component turned out to be significant, we had the most optimal model for 1 cluster. Next, we tested a two cluster model with first two quadratic components and in case of significances only a linear component. This ‘forward procedure’ was repeated for a three, four and five cluster model. Constantly, the Vuong-Lo-Mendell Rubin Likelihood Ratio Test (LMR-LRT), Bootstrapped Likelihood Ratio Test (BLRT) and the Bayesian Information Criterion (BIC) fit indices were compared\textsuperscript{23}. A model indicated better fit when the LMR-LRT and BLRT were significant and the BIC was lower compared to the model with one less cluster\textsuperscript{23}. Internal reliability of the clusters was assessed with the entropy statistics and average posterior probabilities. Statistics >0.80 indicated good classification\textsuperscript{23}. Finally, we looked at the interpretability of the trajectories, the sample size and usefulness of the identified clusters. As LCGM allows missing data, we performed a sensitivity analysis in which we compared the forward procedure in a sample with complete and incomplete PF data. For the final model, we determined the baseline value (intercept=I) and level of change (slope=S) over time for all identified clusters.

Next, we described the characteristics of the identified clusters and performed multinomial regression analysis to examine which indicators were able to discriminate between the identified clusters. Since we found three clusters with two distinct trajectories over time (improvement versus deterioration), we decided to develop a prognostic model for deterioration (belonging to cluster 2-3) versus improvement in PF (belonging to cluster 1). Indicators that were univariately associated (P<0.10) with the outcome were entered into multivariable regression analysis, in which manual backward selection procedure was carried out (P\textsubscript{removal}=0.05) to obtain a final model with prognostic indicators for deterioration in PF (cluster 2-3). The performance of the model was tested, by calculating the area under the curve (AUC) to assess discrimination, Nagelkerke’s pseudo R\textsuperscript{2} to assess the explained variance and the Hosmer-Lemeshow goodness-of-fit test to assess calibration\textsuperscript{24-27}. The internal validity
of the model was tested by using bootstrapping techniques (200 samples), to test for possible overoptimism of the model\textsuperscript{26,27}.

LCGM was performed in M-plus version 7.11, multinominal regression analysis in IBM SPSS version 20.0 and the development and validation of the prediction model in R, version 2.15.

RESULTS

We selected 3687 potential patients in the GPs based on age and the presence of two or more chronic diseases, of which 2706 received a screening questionnaire for joint pain. Of the 2706 potential participants, 1027 were not eligible (no joint pain on most days), 864 did not respond to the screening questionnaire, 106 appeared to suffer from other pain complaints than joint pain (cramps, muscle pains, neurological pains) and 19 were excluded because of other reasons. Of the 690 eligible participants eventually 407 were included. The other 283 participants were excluded because the participants provided no consent for further contact (n=173), we were unable to make an appointment (n=71) or because participants dropped-out during the baseline measure (n=39). This latter drop-out was due to several reasons, for example because of the burden of the tests and questionnaire or motivational problems. More details about this recruitment and exclusion of participant have been recently published \cite{11}.

Within the group of eligible patients, we found no differences between participants (n=407) and non-participants (n=283) in age and gender. However, the non-participants had fewer chronic diseases (mean: 2.59, SD: 0.9 vs. mean: 2.74, SD: 0.9), reported fewer joint pain sites (mean: 3.67, SD: 2.1 vs. mean: 4.19, SD 2.1) and lower pain severity (mean: 64.9, SD: 18.6 vs. mean: 68.4, SD: 15.6).

Of the included 407 participants, 317 completed the study (77.9%; 2 had missing data at T3). There was some drop-out: 9.1% after 6 months, 7.8% after 12 months and 6.5% after 18 months. Most important reasons were death and deteriorated health (Figure 1). More specifically, data on PF were sometimes missing. Compared to the group with complete PF data (75%), the incomplete group (25%) was older, lower educated, more often frail and had more depressive symptoms.

Based on the fit indices, internal reliability and interpretability, the most optimal model was a three-cluster model, as the two-cluster model showed worse fit indices and the four-cluster model showed a non-significant LMR-LRT, as compared to the three-cluster model. Also, the entropy and average posterior probabilities were better in the three-cluster model (Table 1). Additionally, the interpretation of the three-cluster model seemed more relevant, since two clusters from the four-cluster model with similar slopes (-.937 and -1.317 respectively) were combined in one cluster in the three-cluster solution. The sensitivity analysis identified the
same clusters in the sample with incomplete data. Therefore, further analyses were performed in the complete sample. The three identified trajectories are illustrated in Figure 2.

**Figure 1 | Flowchart of study population and dropouts**
<table>
<thead>
<tr>
<th></th>
<th>BIC</th>
<th>BLRT</th>
<th>LMR-LRT</th>
<th>Entropy</th>
<th>N</th>
<th>Posterior probability</th>
<th>Intercept</th>
<th>Slope linear</th>
<th>quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12496.001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.883</td>
<td>209</td>
<td>0.970</td>
<td>70.141</td>
<td>2.553</td>
<td>-0.851</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>198</td>
<td>0.964</td>
<td>28.836</td>
<td>-0.935</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>12157.515</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.868</td>
<td>140</td>
<td>0.913</td>
<td>75.64</td>
<td>3.44</td>
<td>-1.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td>0.950</td>
<td>49.59</td>
<td>-1.28</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>137</td>
<td>0.960</td>
<td>20.70</td>
<td>-0.94</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>12045.330</td>
<td>&lt;0.0001</td>
<td>&lt;0.3926</td>
<td>0.835</td>
<td>93</td>
<td>0.925</td>
<td>79.91</td>
<td>5.07</td>
<td>-1.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>114</td>
<td>0.883</td>
<td>62.12</td>
<td>-0.94</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>104</td>
<td>0.878</td>
<td>39.50</td>
<td>-1.32</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>0.930</td>
<td>16.11</td>
<td>-0.86</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>11982.881</td>
<td>&lt;0.0001</td>
<td>0.0050</td>
<td>0.817</td>
<td>83</td>
<td>0.872</td>
<td>82.88</td>
<td>4.97</td>
<td>-1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>92</td>
<td>0.856</td>
<td>68.24</td>
<td>-0.03</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63</td>
<td>0.904</td>
<td>51.56</td>
<td>-1.74</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
<td>0.843</td>
<td>32.83</td>
<td>-1.42</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78</td>
<td>0.907</td>
<td>13.82</td>
<td>-0.83</td>
<td>*</td>
</tr>
</tbody>
</table>

**Abbreviations:** BIC, Bayesian Information Criterion; BLRT, Bootstrapped Likelihood Ratio Test; LMR-LRT, Vuong-Lo-Mendell-Rubin Likelihood Ratio Test.
The first cluster was defined as ‘good PF’, as the 140 participants showed high baseline PF (intercept: 75.6; 95% CI 72.7 to 78.5) and some improvement over time (slope: 3.4; 95% CI 0.2 to 6.7). Cluster 2 was defined as ‘moderate PF’, as the 130 participants showed average baseline PF (intercept: 49.6; 95% CI 45.0 to 54.1) and some deterioration over time (slope: -1.3; 95% CI -0.1 to -2.4). Cluster 3 was defined as ‘poor PF’, as the 137 participants showed low baseline PF (intercept: 20.7; 95% CI 17.7 to 23.7) and some deterioration over time (slope: -0.9, 95% CI -0.1 to -1.7). Overall, it seemed that especially the baseline PF values (intercepts) differed between the three clusters. There were only small changes in PF over time, which showed either trajectories of improvement (cluster 1) or trajectories of deterioration (cluster 2-3).

![Estimated trajectories of the three clusters](image)

Figure 2 | Three identified clusters and their individual trajectories over time, as measured by the RAND-36 domain physical functioning (PF): absolute change scores.

The characteristics of the participants in the three identified clusters are described in Table 2. Of the 407 included participants, 395 (97%) had complete baseline data of the prognostic indicators. The group comparison revealed that participants with moderate and poor PF over time, were older, had more chronic diseases, reported more depressive symptoms, less self-efficacy and more activity avoidance, compared to the group with good PF. On top of these results, being frail and having pain in more joints were extra indicative of showing poor PF, compared to good PF (Table 3).
Since we found three clusters with two distinct trajectories over time (improvement versus deterioration), we developed a prognostic model for deterioration (belonging to cluster 2-3) versus improvement in PF (belonging to cluster 1). All prognostic indicators were entered in the model. The final model included the following prognostic indicators: higher age, more depressive symptoms, less perceived self-efficacy and more activity avoidance (Table 4). The explained variance of the model was 59%, indicating that most important indicators were included in the model. The AUC of 0.90 reflected good discrimination between participants with improvement versus deterioration in PF. The Hosmer-Lemeshow test was not significant, which indicated acceptable calibration. The shrinkage factor, i.e. level of overfitting of the model, was 0.86.
### Table 3 | Multinomial regression analysis to distinguish between the three identified clusters (Cluster 1: good PF n=138 was used as reference group)

<table>
<thead>
<tr>
<th></th>
<th>Cluster 2 Moderate PF (n=125)</th>
<th>Cluster 3 Poor PF (n=132)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age</td>
<td>1.10**</td>
<td>1.04-1.17</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>1.09</td>
<td>0.56-2.14</td>
</tr>
<tr>
<td>Living situation (alone)</td>
<td>1.47</td>
<td>0.75-2.91</td>
</tr>
<tr>
<td>Education (primary)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>0.91</td>
<td>0.44-1.88</td>
</tr>
<tr>
<td>College/university</td>
<td>0.43</td>
<td>0.17-1.09</td>
</tr>
<tr>
<td>Number of chronic diseases (≥3)</td>
<td>1.88*</td>
<td>1.02-3.46</td>
</tr>
<tr>
<td>Number of joint pain sites (1-8)</td>
<td>1.11</td>
<td>0.93-1.32</td>
</tr>
<tr>
<td>Pain intensity (0-100)</td>
<td>1.00</td>
<td>0.98-1.02</td>
</tr>
<tr>
<td>Frailty (yes)</td>
<td>0.91</td>
<td>0.22-3.75</td>
</tr>
<tr>
<td>Depressive symptoms (0-21)</td>
<td>1.11*</td>
<td>1.00-1.24</td>
</tr>
<tr>
<td>Self-efficacy (range 6-60)</td>
<td>0.93**</td>
<td>0.90-0.96</td>
</tr>
<tr>
<td>Avoidance of activities (range 5-20)</td>
<td>1.16**</td>
<td>1.05-1.29</td>
</tr>
<tr>
<td>Catastrophizing (range 0-6)</td>
<td>1.19</td>
<td>0.95-1.50</td>
</tr>
<tr>
<td>Social support (range 12-60)</td>
<td>1.01</td>
<td>0.97-1.05</td>
</tr>
</tbody>
</table>

**Abbreviations:** PF, physical functioning; OR, odds ratio

**P<0.01 | P<0.05**

### Table 4 | Final model of prognostic indicators for moderate-poor physical functioning (reference group: good physical functioning), multivariable logistic regression analysis with backward selection

<table>
<thead>
<tr>
<th>Prognostic indicators</th>
<th>moderate - poor PF (n=257)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
</tr>
<tr>
<td>Age</td>
<td>1.10</td>
</tr>
<tr>
<td>Depressive symptoms (0-21)</td>
<td>1.18</td>
</tr>
<tr>
<td>Self-efficacy (range 6-60)</td>
<td>0.91</td>
</tr>
<tr>
<td>Avoidance of activities (range 5-20)</td>
<td>1.20</td>
</tr>
</tbody>
</table>

**Area under the Curve (discrimination): Original 0.90 – Shrunk 0.88**

**Nagelkerke’s R2: Original 0.59 – Shrunk 0.53**

**Hosmer and lemeshow test: 0.79**

**Slope (calibration): 0.86**

**Abbreviations:** OR, odds ratio
DISCUSSION

Summary of findings
In our observational cohort study, we found that older adults with joint pain and comorbidity only showed modest changes in PF over a period of 18 months. However, latent class growth modelling revealed more underlying variation in PF and identified three homogeneous subgroups with two distinct trajectories of PF. Participants in cluster 1 showed good baseline PF and some improvement over time, whereas cluster 2 and 3 showed moderate and poor baseline PF respectively and some deterioration over time. The classification of the clusters seemed to depend mostly on baseline status of PF and to a lesser extent on the slopes of the trajectories of participants. The final prognostic model for improvement (belonging to cluster 1) versus deterioration in PF (belonging to cluster 2-3) included the following indicators: higher age, more depressive symptoms, less perceived self-efficacy and more activity avoidance. This model showed good calibration and discrimination, even after adjusting for overoptimism.

It is well described in the literature that it remains complicated to compare models based on model fits alone, as the model fit indices often deviate and thus are not in agreement with each other\textsuperscript{23}. Our results showed that the BLRT and LMR-LRT did not point to one definite solution. Hence, we took the clinical interpretation of the trajectories into account and decided based on the most appropriate solution, as described in the results section. Whether this is the most optimal procedure is up to debate. Nonetheless, the classification of the participants into the three clusters was supported by the differences that we found in the characteristics of the subgroups and the high average probabilities of belonging to the specific subgroups. Participants in cluster 1 already had good levels of PF and improved slightly over time. In contrast many participants in our sample deteriorated in PF, and this decline was present at two levels based on their baseline status. These trajectories may be explained by the ‘horse-racing’ effect\textsuperscript{28}. Presumably, at the point of baseline assessment most patients already had a history of pain complaints and systematic long-term differences in trajectories of PF. Participants who have shown faster deteriorating than average in the past, will continue to deteriorate faster than average in the future. They will drift further away from the rest and experience even more deterioration. The same holds for improvement. Therefore, the participants with higher baseline values will improve further and the participants with lower baseline values will deteriorate further in PF over time. However, we should bear in mind that the observed changes in our study were only minor, as illustrated by the actual change scores in PF (ΔPF) between baseline and 18 months. On a scale of 0-100, the ΔPF was 0.21 (SD 14.6) for cluster 1, -3.9 (SD 18.5) for cluster 2 and -3.0 (SD 11.2) for cluster 3. This indicates fairly stable PF in our sample over 18 months’ time.
Comparison with similar studies
Although this is the first study to report trajectories of PF in older adults with joint pain and comorbidity, our results are similar to those trajectories reported for PF in early symptomatic knee osteoarthritis (OA)\(^2\). However, the model that derived from this study retained mainly physical prognostic indicators, like signs and symptoms of OA (e.g. stiffness, tenderness) and comorbidity, whereas our model showed that it were more the psychosocial factors that were prognostic for poor PF. This discrepancy could be due to the selection of prognostic indicators; on the other hand, participants in the present study were older and psychosocial factors could be a stronger factor at higher ages. The latter could be strengthened by the previous findings that amongst others depression and self-efficacy were indeed important prognostic indicators for poor PF in older adults with musculoskeletal pain\(^5,30,33\). So, our results are in line with these studies, despite the use of different techniques to study PF over time.

Implications
The results of this study have implications for primary care clinicians. We untangled a highly heterogeneous group, by identifying underlying groups with more distinct trajectories of PF. Based on these findings, we developed a model that is intended to help clinicians to understand in more detail how these groups differ from each other in terms of characteristics and how their prognosis can be estimated more accurately. Since especially the psychosocial factors were indicative for poor prognosis of PF, clinicians should pay attention to the presence of such factors in older adults with joint pain and comorbidity, narrow the focus and identify those older adults at risk of poor PF. These predictive factors may also serve as potential targets for interventions aiming to prevent deterioration in physical functioning in older adults with joint pain and comorbidity, but other study designs (like intervention studies) are necessary to further explore such hypotheses.

Strengths and limitations
To our knowledge, no previous studies have analyzed the prognosis of PF in a sample of older adults with joint pain and comorbidity, while the co-existence of joint pain and other chronic diseases could have additional negative effects on prognosis. We focused on this highly relevant group, untangled the population and provided prognostic information on PF. Furthermore, we measured the outcome PF over 18 months at 6 month time-intervals, which enabled the use of LCGM for the identification of underlying trajectories in our heterogeneous population. Also, we included psychosocial factors as possible prognostic indicators, besides the widely accepted and included health-related factors. Finally, we used valid and reliable instruments to measure our outcome and prognostic indicators. We should also mention some limitations. The significant differences between the participants and non-participants in the eligible group
indicated some selection bias, but we argue that this selection bias has only limited impact on the generalizability of the study results as the actual level of differences appeared to be only small (mean and SD). Although we had a good response rate during follow-up (78%), some drop-out occurred. However, the sensitivity analysis showed similar clusters with LCGM in the sample with complete and incomplete data, which indicates that bias because of missing data is highly unlikely. Furthermore, 97% of the data on prognostic indicators was complete, because we visited all participants at their homes during the baseline measure, which enabled us to check if the questionnaire was completed. Therefore, our final prognostic model seems valid for the entire study population. There may be some problems with the applicability of our results in primary care, as participants of 65 years or older with comorbidity were further included based on self-reported pain and not based on actual consultation data for joint pain. However, supplementary analysis showed that 363 (90%) of our participants visited their GP for their joint pain complaints, of which 251 (69%) of the consultations took place in the year prior to inclusion. This indicates that general practitioners are aware of most reported pain complaints. Unfortunately, we had no data on actual diagnosis of the musculoskeletal complaints (e.g. osteoarthritis, rheumatic diseases) to perform subgroup analysis. One could argue that participants with similar types of arthritis are grouped into similar clusters. For example, symptoms of participants with osteoarthritis (OA) may differ from symptoms of participants with rheumatoid arthritis (RA), which could result in one cluster with largely OA participants, and another cluster with RA participants. As we do not have information on the specific type of arthritis, we can only speculate about grouping of specific types of arthritis into specific clusters in our analysis. Furthermore, one could argue about the length of our follow-up period. Many former studies showed fairly stable courses of PF over 3-5 years. However, these studies assumed one single trajectory representing all individuals in the dataset. It was unknown whether these findings would hold in a study that assumed underlying subgroups.

CONCLUSION

Participants with joint pain and comorbidity either showed good baseline PF with some improvement (cluster 1) or moderate-poor baseline PF with some deterioration over time (cluster 2-3). Most important prognostic indicators of belonging to cluster 2-3, compared to belonging to cluster 1 were higher age, more depressive symptoms, less perceived self-efficacy and more avoidance of activities. The prognostic model facilitates the classification of patients, the provision of more accurate information about prognosis and helps to narrow the focus to the high risk group of poor PF.
REFERENCES


