Sequence Analysis to Phenotype Healthcare Patterns in Adults with Musculoskeletal Conditions Using Primary Care Electronic Health Records.

MATHEW, Smitha, PEAT, George <http://orcid.org/0000-0002-9008-0184>, PARRY, Emma, WILKIE, Ross, JORDAN, Kelvin P, HILL, Jonathan C and YU, Dahai <http://orcid.org/0000-0002-8449-7725>

Running head: Healthcare patterns of MSK conditions in primary care

Title: Sequence Analysis to Phenotype Healthcare Patterns in Adults with Musculoskeletal Conditions Using Primary Care Electronic Health Records

Smitha Mathew, MSc[1], George Peat, MSc, PhD [1,2], Emma Parry, MRCGP, PhD [1], Ross Wilkie, BSc, PhD[1], Kelvin P Jordan, MSc, PhD[1], Jonathan C Hill, MSc, PhD[1], Dahai Yu, MSc, PhD[1]

Affiliations

[1] Keele University, Staffordshire, UK

[2] Sheffield Hallam University, Sheffield, UK

Corresponding author details:

Dr. Dahai Yu,

Primary Care Centre Versus Arthritis**,** School of Medicine,

Keele University, Staffordshire, UK ST5 5BG

Tel.: +44-0-1782-734891; Fax: +44-0-1782-734719

Email: d.yu@keele.ac.uk

ORCiD: Dahai Yu (https://orcid.org/0000-0002-8449-7725)

Funding

The MIDAS-GP study has been funded by the Nuffield Foundation's Oliver Bird Fund and Versus Arthritis (OBF/43390), but the views expressed are those of the authors and not necessarily the funders. Visit www.nuffieldfoundation.org and www.versusarthritis.org. The MIDAS-GP Study is supported by Keele Clinical Trials Unit and registered on the ISRCTN Registry (ISRCTN18132064). EP is funded by a National Institute for Health and Care Research (NIHR) Academic Clinical Lectureship CL-2020-10-001. KPJ is partly funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration West Midlands. The views

expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Article type: Original article

Word count: 3,846 words, 4 Tables, 2 Figures, 34 references

Abstract

Objective

The aim of this study was to apply sequence analysis (SA) to phenotype healthcare patterns of adult patients with musculoskeletal (MSK) conditions using primary care electronic health records and to investigate the association between these healthcare patterns and post-consultation patient's self-reported outcome.

Methods

Data from the Multi-level Integrated Data for musculoskeletal health intelligence and ActionS (MIDAS) programme conducted in North Staffordshire and Stoke-on-Trent, UK was utilised. The study included patients aged ≥18 years who consulted primary care for MSK conditions between September 2021 and July 2022. SA was employed to categorise patients with similar healthcare patterns in primary care in the five years prior to their index consultation in respect to consultations, analgesic prescriptions, imaging, physiotherapy, and secondary care referrals. Association of socio-demographic characteristics and self-reported outcome with clusters were determined.

Results

In total, 1,875 patients consulting primary care for MSK conditions were available for analysis. SA identified five clusters of prior healthcare patterns among patients with MSK conditions, including "increasing consultation and analgesia" (5.60%), "low consultation and healthcare use" (57.39%), "high consultation and healthcare use" (8.32%), "low consultation but high analgesia" (13.01%), and "low consultation but moderate healthcare use" (15.68%). Patients in the "high consultation and healthcare use" group were predominantly female, older, obese, had more comorbidities and lived in the most deprived areas compared to those in the "low consultation and

healthcare use" group. Additionally, self-reported outcome varied significantly between clusters, with patients in the "high consultation and healthcare use" group reporting worse self- reported outcome.

Conclusion

This analysis identified five distinct clusters of healthcare patterns for patients with MSK conditions in primary care and observed substantial variations in patient's self-reported outcome and socio-demographic profiles across these different groups of patients.

Key words: sequence analysis, optimal matching, cluster analysis, healthcare patterns, musculoskeletal conditions, primary care, electronic health records

Significance and Innovations

- Our study identified five distinct patterns of healthcare utilisation in primary care among adult patients with MSK conditions using sequence analysis.
- We observed inequalities in healthcare utilisation patterns based on patients' characteristics and significant variations in patients' self-reported outcome across different clusters of healthcare utilisation patterns. Specifically, patients from socio-economically deprived areas, who were predominantly older, female patients, obese, and had multiple comorbidities showed higher consultation rates, healthcare use and poorer short-term outcome.
- These findings highlighted the importance of addressing disparities in healthcare access and the need for targeted interventions for patients at risk of poorer health outcomes.

Introduction

Musculoskeletal (MSK) conditions are a major cause of pain and disability worldwide. In the UK, more than 20 million people live with a MSK condition [1]. MSK conditions are primarily assessed and managed in primary care. It accounts for 12–14% of primary care consultations in adults, and a substantial portion of healthcare expenditure is allocated to managing these conditions [2]. A range of different interventions are recommended for the management of MSK conditions including providing advice on self-management and exercise, referring patients for non-pharmacological treatments like physiotherapy, and prescribing analgesics to alleviate pain and symptoms [3].

Pain associated with MSK conditions leads to high healthcare use, and patients seeking healthcare may find themselves consulting a diverse range of healthcare professionals and receiving a mix of analgesic prescriptions, imaging, physiotherapy and secondary care referrals [4,5]. Understanding patterns within these interactions can provide insights into how different patient subgroups utilise healthcare services. By analysing these patterns, healthcare providers can identify the specific needs of patient subgroups. It enables healthcare planners to allocate resources more strategically, ensuring that they are directed to where they are most needed. Moreover, a comprehensive understanding of care patterns helps identify service gaps and areas for improvement. This knowledge allows for the optimisation of healthcare delivery by addressing disparities in service utilisation. Overall, it supports identifying specific healthcare needs, informs strategic resource allocation, and contributes to improving healthcare delivery and patient outcomes [6,7].

Patient' self-reported outcome measures are valuable for evaluating perceptions of health, symptoms, and the effectiveness of MSK management [8]. These measures capture information primarily focusing on pain levels, activity limitations, and overall quality of life rather than clinical measures [9]. Several studies have highlighted an association between chronic pain and increased

healthcare utilisation [10–12]. Additionally, a correlation has been observed between low health-related quality of life and high healthcare utilisation [12]. Evidence from a primary care prospective observational cohort study further indicates that subgroups of individuals with different levels of risk for poor MSK pain outcomes exhibit different levels of healthcare utilisation [13]. Relating healthcare utilisation patterns to patient' self-reported outcome might direct attention towards potentially poorly targeted or ineffective patterns of care.

In recent years, sequence analysis (SA) has emerged as a promising analytical approach in healthcare research due to its ability to uncover valuable insights and patterns from real-world data [14]. SA is used to analyse ordered sets of data, often referred to as sequences. This method is commonly used in social science to identify patterns in life course trajectories and to study transitions into adulthood [15,16] or career patterns [17] by examining longitudinal data representing events experienced by individuals over time. In healthcare, SA allows researchers to analyse sequences of medical events, such as diagnoses, treatments, and procedures, to understand disease progression and care pathways [6,18,19]. SA enables the exploration of healthcare utilisation patterns, including patient journeys through the healthcare system, patterns of service utilisation, and transitions between different levels of care [20–22].

A conventional SA involves three steps: defining events as sequences of successive categorical states, calculating dissimilarities between pairs of sequences, and building a typology of the sequences [23]. The states in the sequence should be clinically meaningful and relevant to the research objectives. Dissimilarity is a quantitative measure indicating the degree to which two individuals followed distinct sequences. There are different dissimilarity measures based on alignment and non-alignment techniques. The choice of dissimilarity measure may impact the

results of SA; therefore, researchers select an appropriate measure aligned with their research objective [24]. Finally, a cluster analysis is performed to classify individuals with similar sequences. In this study, we focus on the identification of different healthcare patterns among adult patients with MSK conditions in primary care over five years prior to their index consultation, as well as examining the impact of these patterns on patient's self-reported outcome. By examining historical care patterns, we can comprehensively understand the various treatment strategies patients have experienced, which might influence their current health status and outcome.

Therefore, the primary objective of this study was to apply SA to phenotype healthcare patterns of patients with MSK conditions from routinely collected primary care electronic health records (EHRs). The secondary objective was to investigate the association between the identified healthcare patterns and post-consultation patient's self-reported outcome.

Materials and methods

Data source and population

The Multi-level Integrated Data for musculoskeletal health intelligence and ActionS (MIDAS) programme, funded by the Nuffield Foundation and Versus Arthritis, aims to develop a comprehensive, place-based system for MSK health data in North Staffordshire and Stoke-on-Trent, UK. MIDAS-GP is one observational cohort study within the overall MIDAS project, and is designed to collect, link, and explore data from patient-report, electronic health records, and other sources for adults presenting with common, painful MSK conditions presenting in general practice. The study focuses on integrating data from various clinical settings to enhance MSK care pathways. The pre-specified MIDAS-GP study protocol is available at Open Science Framework (https://osf.io/e542w/). The study received ethical approval from Yorkshire & The Humber - Leeds West Research Ethics Committee (Reference: 21/YH/0178). The eligible participants for this study

included patients aged 18 years and older, registered with thirty participating general practices and who consulted any primary care healthcare professional within the practice for a painful, non-inflammatory MSK condition. Recruitment was conducted from September 2021 to July 2022, staggered across different practices, with recruitment periods lasting from three to six months. Relevant MSK pain-related consultations were identified using a pre-specified Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) code list (Supplementary Table S9). Eligible participants were invited to complete a baseline questionnaire on MSK health and care and asked for their consent to linkage to electronic health records. The consenting participants were further asked to complete the follow-up questionnaires at 3- and 6- months. The information on patient's demographic, socioeconomic, comorbidities, and MSK management strategies were derived from the primary care EHR in the five years prior to index MSK consultation. The list of comorbidities used was produced after cross-mapping morbidities in National Institute for Health and Care Excellence (NICE) multimorbidity indicator for general practice [25], Charlson [26] and Elixhauser [27] comorbidity indices, and potentially relevant case-mix adjustment methods [28]. Comorbidity code lists are available at Open Science Framework (https://osf.io/e542w/). The MSK management information included MSK-related primary care consultations, relevant prescriptions for medications, referrals for imaging (e.g. radiographs, magnetic resonance imaging (MRI), or CT scans), referrals for physiotherapy, and referrals for secondary care (MSK triage, rheumatology, trauma and orthopaedic departments). Patients' neighbourhood deprivation was also considered. We used the English index of multiple deprivation (IMD) 2019 rank as a composite measure of neighbourhood deprivation, which covers seven domains of material deprivation including income, employment, education and skills training, health deprivation and disability, barriers to housing and services, crime, and living environment [29]. The IMD classifies

the areas into five quintiles based on relative disadvantage, with quintile 1 being the most deprived and quintile 5 being the least deprived. Additionally, patients' MSK-Health Questionnaire (HQ) score at baseline, 3-months, and 6-months after index consultation were considered. MSK-HQ score is a 14-item questionnaire that captures key outcomes that patients with MSK conditions have prioritised as important for use across clinical pathways [30]. Scores range from 0 to 56, with higher scores indicating better MSK health over the past two weeks [30]. The data of this study is available upon request.

Statistical analysis

To explore the patterns of utilisation of key MSK management strategies in primary care, we employed a multichannel SA involving five domains: MSK consultations, analgesic prescriptions, imaging referrals, physiotherapy referrals, and secondary care referrals. The primary step in SA was defining the states within the sequence, the observation period, and the time unit. The healthcare patterns of patients with MSK conditions were observed for five years before their index consultation. The MSK management information was retrieved as annual count data. So, we defined three categorical states for consultations and analgesic prescriptions: "None", "Low", and "High", representing 0, 1-3, and 4 or more instances, respectively and two categorical states for imaging, physiotherapy, and secondary care referrals: "No" and "Yes" occurrence during the year (detailed in the Supplementary File). If the care event is not recorded in the system, it is considered to have not occurred. We defined care sequences for each domain for each patient, with each sequence consisting of five states (one for each year).

For the analysis of sequences, we chose optimal matching (OM) edit distance, the most often used approach to measure the dissimilarity between pairs of sequences [14]. OM measures the dissimilarity between two sequences by determining the minimum cost required to transform one

sequence into another by edit operations such as insertion, deletion, or substitution of states. We opted for a data-driven cost for indels (insertion and deletion), and substitutions based on the frequency of the states in the sequences, referred to as INDELSLOG. In this approach, indel costs were calculated initially as the logarithm of the inverse of the relative frequency of the states, as log[2/(1 +f)], where f is the relative frequency of the states. Then, the substitution costs between the two states are computed by summing their indel costs [31]. The rationale behind this approach is that inserting or deleting rare states is more costly than inserting or deleting frequent states and substituting rarely observed states costs more than substituting common states [31]. The multidomain dissimilarity matrix was computed by adding the domain-specific dissimilarity matrices. Based on the computed dissimilarity matrix, we performed an agglomerative hierarchical cluster analysis with Ward's linkage to classify patients with similar care patterns. The optimum number of clusters was determined based on the dendrogram, inertia jump curve, cluster quality indices, and clinical relevance and interpretability (explanation of the selection criteria is given in the Supplementary File). To visualise the care patterns, we used sequence index plots and state distribution plots provided by the SA. State distribution plot shows the distribution of states for each time unit, while each line in the sequence index plot represents an individual sequence [24]

We compared patients' demographic and health characteristics between the derived clusters using the Chi-square test, t-test, and ANOVA. A multinomial logistic regression model was used to assess the association between patient's profiles and cluster membership. A linear mixed model was used to test the difference in patient-reported MSK-HQ score between clusters. The model included the fixed, categorical effects of cluster, time, cluster-by-time interaction, gender, comorbidities, and IMD, alongside continuous, fixed covariates for age and BMI. To account for within-subject variability, an unstructured covariance structure was applied to model the within-

subject errors. The missing data in BMI (n=280) was imputed by multiple imputation using chained equations [32]. Sensitivity analyses were conducted to ensure the reproducibility of the results. For this, MSK patients were sub-grouped into those with osteoarthritis (OA) and those with low back pain (LBP), and the SA was repeated within these sub-groups. The SA was carried out using the TraMineR and WeightedCluster packages in R and all other analyses were performed using STATA 18.

Results

Participants

A total of 2008 (14.9%) patients responded at baseline, of which 1875 consented and were successfully linked to their EHR and hence form the primary population for analysis (detailed in Supplementary Figure S1). Among these patients, the mean (SD) age was 57.74 (15.50) years, and the mean (SD) BMI was 29.18 (6.91) kg/m^2. Female subjects accounted for 65.76% of the patients, while 32.43% were classified as obese, and 28.27% were from the most deprived areas (Table 1). Patients' care sequences of each domain were presented in sequence index plots (Supplementary Figure S2).

Sequence analysis

By the multichannel SA of the domains—MSK consultations, analgesic prescriptions, imaging referrals, physiotherapy referrals, and secondary care referrals—patients with similar care sequences were classified into five distinct clusters (Figure 1) based on the dendrogram, inertia jump curve and cluster quality indices (Supplementary Figures S3 & S4 and Table S1). The characteristics of the identified clusters are as follows:

- Cluster 1 (n=105, 5.60%) patients were characterised by a marked increase in high-level (i.e. 4 or more) consultations and analgesic prescriptions over the five years, accompanied

by moderate imaging and physiotherapy, and minimum secondary care referrals. This cluster can be labelled as “increasing consultation and analgesia”.

- Cluster 2 (n=1076, 57. 39%) consisted of patients with low-level (1-3) consultations and analgesic prescriptions mainly in the index year, and minimal imaging, physiotherapy, and secondary care referrals. This cluster can be labelled as “low consultation and healthcare use”.
- Cluster 3 (n=156, 8.32%) was made up of patients with consistently higher levels of consultation, analgesic prescriptions, imaging, physiotherapy, and secondary care referrals. This cluster can be labelled as “high consultation and healthcare use”.
- Cluster 4 (n=244, 13.01%) included patients with low-level (1-3) consultations, low imaging, physiotherapy, and secondary care referrals, but having higher levels (4 or more) of analgesic prescriptions over the five years. This cluster can be labelled as “low consultation but high analgesia”.
- Cluster 5 (n=294, 15.68%) consisted of patients with low-level (1-3) of consultations, analgesic prescriptions and secondary care referrals, but moderate levels of imaging and physiotherapy referrals. This cluster can be labelled as “low consultation but moderate healthcare use”.

Potential predictors of cluster membership

Patients’ characteristics by clusters of similar care patterns were presented in the Supplementary Table S2. Table 2 shows the findings of the multinomial logistic regression model computed to examine potential predictors of the identified clusters. Odds ratios (OR) were calculated to indicate the likelihood of being in a particular cluster compared to the reference cluster. The reference cluster used in the analysis was “low consultation and healthcare use”. Female patients were

significantly more likely to be in clusters "high consultation and healthcare use", "low consultation but high analgesia", and "low consultation but moderate healthcare use", as compared to being in "low consultation and healthcare use" cluster. Additionally, older age, obesity, a higher comorbidity index, and socio-economic deprivation (most deprived) were identified as significant predictors for membership in clusters "increasing consultation and analgesia", "high consultation and healthcare use", "low consultation but high analgesia", and "low consultation but moderate healthcare use".

The impact of healthcare patterns and patients' MSK-HQ score

Table 3 presents the adjusted estimates for the association between healthcare patterns and MSK-HQ score. Figure 2 illustrates the predicted mean MSK-HQ score values among different clusters at index consultation (baseline), and at 3-months and 6-months following the index consultation. The mean patient-reported MSK-HQ score was significantly lower (worse MSK health) in clusters "increasing consultation and analgesia", "high consultation and healthcare use", "low consultation but high analgesia", and "low consultation but moderate healthcare use" compared to the cluster "low consultation and healthcare use" at baseline, 3-months, and 6-months; the estimated differences in mean score are presented in Table 4. Additionally, the MSK-HQ score over time, as indicated by the interaction terms of clusters with similar care sequences and time, showed significantly less improvement at month 3 in "high consultation and healthcare use" (-5.18 [95% CI: -6.92, -3.43]), "low consultation but high analgesia" (-2.88 [-4.28, -1.49]), and "low consultation but moderate healthcare use" (-1.93 [-3.30, -0.57]) compared to the improvement in "low consultation and healthcare use". Similarly, less improvement was observed at month 6 in "increasing consultation and analgesia" (-2.82 [-5.27, -0.37]), "high consultation and healthcare use" (-4.57 [-6.55, -2.58]), "low consultation but high analgesia" (-2.98 [-4.61, -1.35]) (Table 3).

Sensitivity analysis

To test the reproducibility of the results, two additional SA were performed by subgrouping the MSK patients into those with OA and those with LBP. Agglomerative hierarchical cluster analysis with OM and INDELSLOG cost produced five clusters for patients with OA (Supplementary Figure S5), which were similar to the results obtained for patients with MSK conditions. Similarly, the analysis of patients with LBP also resulted in five clusters (Supplementary Figure S6). Another sensitivity analysis with similar SA methods was conducted excluding patients who have <5 years of continuous retrospective record, and it yielded similar clusters of the main analysis (Supplementary Figure S9).

Discussion

Our study examined healthcare patterns among 1,875 adult patients who sought consultation for MSK conditions in primary care settings and investigated the relationship between these healthcare patterns and the patient's self-reported MSK-HQ outcome. Using SA, we identified five distinct clusters that differed in terms of MSK-related pain consultations, analgesic prescriptions, imaging, physiotherapy, and secondary care referrals. The data tells us that the low consultation and healthcare use group have the best MSK health. Factors associated with being in the other clusters and poorer health are gender, age, BMI, comorbidities and neighbourhood deprivation.

To our knowledge this is the first study to use SA methodology to uncover healthcare patterns of MSK conditions in primary care using routinely collected EHR data. A Canadian study by Nguefack et al. used SA to identify five 2-year care trajectories among patients living with arthritic conditions [6]. However, their focus was on patterns of healthcare visits across different healthcare services (e.g., emergency department visits, hospitalisations and pain clinics) without considering multiple treatment strategies. This may be due to variations in the healthcare systems, which may

influence the applicability of different primary care approaches. Similarly, Mose et al. employed latent class growth analysis to identify five 10-year patterns of MSK healthcare utilisation among adult Danes who reporting chronic MSK pain [5]. While they modelled the number of healthcare contacts, they did not analyse the sequence of services used. Our findings have similarities with trajectories from studies analysing single components of healthcare. However, in contrast our study examined jointly all the main components of MSK management in primary care settings. Additionally, Meisingset et al. identified five distinct MSK phenotypes using latent class analysis, but their focus was on key prognostic factors over the biopsychosocial domains across common MSK pain [33]. While these phenotypes may support the development of targeted interventions, our study, which integrates different care strategies for MSK pain in primary care, offers practical insights that may enhance clinical practice and inform decision-making in primary care settings.

This study demonstrated that patients in the "high consultation and healthcare use" group experienced the worst outcome in terms of MSK-HQ score. This finding aligns with the results of the Nguefack et al. study, which indicated that belonging to a high healthcare utilisation group was associated with a higher likelihood of perceiving a poor or fair quality of life [6]. This high-utilisation group in our study represented 8.32% of MSK consulters and predominantly consisted of female subjects, older patients, and obese individuals and those coming from the most deprived areas. Additionally, this group had the highest proportion of patients with a comorbidity count of 3 or above, suggesting a significant burden of comorbidities.[6]

In contrast, patients in the "low consultation and healthcare use" group exhibited the best MSK health (highest MSK-HQ score). This was the largest group, comprising 57.39% of MSK consulters, and included a higher proportion of male subjects, younger patients, fewer obese individuals, and a greater proportion of patients who with no comorbidities. Notably, 389 (36.15%)

patients in this group had consultations only in year one, suggesting they might be incident consulters. Furthermore, individuals from the least deprived areas typically use healthcare services less frequently than those from the most deprived areas, a finding consistent with other studies reporting socio-economic differences in the prevalence and management of chronic pain [34]. These results indicate that more sophisticated sequence analysis nevertheless confirms the general observation made in previous studies of a subset of patients with high levels of pain and disability and high healthcare use, in which issues of quality and effectiveness of care may be more important than simple lack of access to primary care.

By evaluating data from the five years prior to the index consultation, we gained insights into the longitudinal treatment strategies experienced by patients. This helps healthcare providers learn from previous cases, refining treatment guidelines and care strategies based on actual outcomes. Furthermore, our approach helps identify patient groups that require more intensive and tailored care, allowing for a more effective allocation of resources to where they are needed most. Our findings reveal that nearly half of the patients consulting for MSK conditions have a long history of healthcare interactions, which is associated with poorer short-term outcome. These patients typically come from socio-economically deprived areas and are predominantly older, female, and obese, with multiple comorbidities. Our assessment of patients' profiles and outcome variations between healthcare utilisation patterns can be used to improve care pathways and highlights areas where policy interventions could substantially enhance health equity.

The strength of this study lies in its innovative multidimensional approach to SA, enabling a comprehensive exploration of the most shared healthcare utilisation patterns for MSK conditions in primary care, considering patterns of consultations, analgesic prescriptions, imaging, physiotherapy, and secondary care referrals. There are potential limitations in this study. The

inclusion of only those patients who consented to participate might have introduced a selection bias, as evidenced by the poor response rate. Additionally, the IMD data suggests that the sample was less deprived compared to the general population. Consequently, the patterns of healthcare identified here, and their relative frequency, may not reflect those in the target population of all adult MSK pain consulters. In particular, the frequency of low consultation and healthcare use may be over-estimated in our sample given indirect evidence of lower study participation among more deprived patients. Moreover, our analysis was based on continuous retrospective records of 5 years prior to the index consultation. The registration period of the patients was not available in the data, so we were not sure whether the patients with missing healthcare events had no recorded events or were not registered during that period. We checked whether the patients had 5 years of continuous records by computing the difference between the index date and the date of the first recorded event. We found 738 patients had less than 5 years of continuous retrospective records. Excluding these patients does reduce the sample size. Therefore, we conducted a sensitivity analysis excluding these patients, and the full results are provided in the supplementary file (Supplementary Figures S7-S10 and Tables S3-S8).

Optimising primary care and linkage to effective approaches is crucial for reducing the impact of MSK conditions. Understanding the patterns of patients' journeys through various healthcare services contributes to the achievement of this goal. SA could serve as a feasible method for identifying patient interactions with the healthcare system by delineating sequences of care events and identifying distinct healthcare utilisation patterns. This study offers initial insights into patterns of healthcare by MSK consulters to primary care which have been directed by clinicians. Further investigations are warranted to gain a deeper understanding of care patterns for MSK conditions

in primary and secondary care settings and focus on specific MSK subpopulations such as osteoarthritis and low back pain.

In conclusion, this study identified five distinct healthcare patterns among adult patients with MSK conditions using SA. Patient's self-reported outcome and sociodemographic profiles varied across the five clusters. Patients with high healthcare utilisation reported poorer outcome, while those with lower utilisation had better outcome. These findings underscore the association between socio-economic status, extensive healthcare utilisation, and poorer health outcome, emphasising the need for targeted policy interventions to improve health equity and quality of care.

Acknowledgments

This study is based on data from MIDAS GP study. We are extremely grateful for the involvement and contributions of our Patient Advisory Group to this study. We thank James Bailey for his assistance in retrieving data from electronic health records. The authors wish to acknowledge the contributions of Simon Wathall, Gerri Mulcahy and members of the NIHR Clinical Research Network: West Midlands, practice managers and staff at participating practices, staff at MJog by Livi, and to Sarah Lawton, Steff Garvin, Clare Thompson, Jo Smith, Sarah Lewis, Rachael Heath, Jacqui Carter and the administration support staff in Keele CTU contributed to the design and implementation of practice-based patient recruitment methods for MIDAS-GP.

SM is supported by PhD studentship from the Faculty of Medicine and Health Sciences at Keele University.

Author Contributions

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication.

Study conception and design. SM, GP, EP, RW, KPJ, JCH, DY

Acquisition of data. SM, DY

Analysis and interpretation of data. SM, DY

Ethics

Ethical approval was obtained for the MIDAS GP study from Yorkshire & The Humber - Leeds West Research Ethics Committee (Reference: 21/YH/0178).

References

1. Versus Arthritis. *The State of Musculoskeletal Health 2023*.; 2023. Accessed August 2, 2023. https://www.versusarthritis.org/media/25649/versus-arthritis-state-msk-musculoskeletal-health-2023.pdf
2. Yu D, Missen M, Jordan KP, et al. Trends in the Annual Consultation Incidence and Prevalence of Low Back Pain and Osteoarthritis in England from 2000 to 2019: Comparative Estimates from Two Clinical Practice Databases. *Clin Epidemiol*. 2022;Volume 14:179-189. doi:10.2147/CLEP.S337323
3. Welsh VK, Mason KJ, Bailey J, et al. Trends in consultations and prescribing for rheumatic and musculoskeletal diseases: an electronic primary care records study. *British Journal of General Practice*. 2023;73(736):e858-e866. doi:10.3399/BJGP.2022.0648
4. Babatunde OO, Bishop A, Cottrell E, et al. A systematic review and evidence synthesis of non-medical triage, self-referral and direct access services for patients with musculoskeletal pain. *PLoS One*. 2020;15(7):e0235364. doi:10.1371/journal.pone.0235364
5. Mose S, Kent P, Smith A, Andersen JH, Christiansen DH. Trajectories of Musculoskeletal Healthcare Utilization of People with Chronic Musculoskeletal Pain – A Population-Based Cohort Study. *Clin Epidemiol*. 2021;Volume 13:825-843. doi:10.2147/CLEP.S323903
6. Nguena Nguefack HL, Pagé MG, Choinière M, et al. Distinct care trajectories among persons living with arthritic conditions: A two-year state sequence analysis. *Frontiers in Pain Research*. 2022;3. doi:10.3389/fpain.2022.1014793
7. Flothow A, Novelli A, Sundmacher L. Analytical methods for identifying sequences of utilization in health data: a scoping review. *BMC Med Res Methodol*. 2023;23(1):212. doi:10.1186/s12874-023-02019-y
8. Husselbee R, Price J. Implementing and evaluating patient reported outcome measures (MSK-HQ) using electronic patient records in musculoskeletal practice: Analysis of over 11,000 records. *Physiotherapy*. 2022;114:e94-e95. doi:10.1016/j.physio.2021.12.038
9. Hill JC, Thomas E, Hill S, Foster NE, van der Windt DA. Development and Validation of the Keele Musculoskeletal Patient Reported Outcome Measure (MSK-PROM). *PLoS One*. 2015;10(4):e0124557. doi:10.1371/journal.pone.0124557
10. Kinge JM, Knudsen AK, Skirbekk V, Vollset SE. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. *BMC Musculoskelet Disord*. 2015;16(1):75. doi:10.1186/s12891-015-0536-z
11. Häuser W, Wolfe F, Henningsen P, Schmutzer G, Brähler E, Hinz A. Untying chronic pain: prevalence and societal burden of chronic pain stages in the general population - a cross-sectional survey. *BMC Public Health*. 2014;14(1):352. doi:10.1186/1471-2458-14-352
12. Emilson C, Åsenlöf P, Demmelmaier I, Bergman S. Association between health care utilization and musculoskeletal pain. A 21-year follow-up of a population cohort. *Scand J Pain*. 2020;20(3):533-543. doi:10.1515/sjpain-2019-0143

13. Oppong R, Lewis M, Campbell P, et al. Comparison of health-care utilization, costs and health-related quality of life across the subgroups defined by the Keele STarT MSK Tool. *Rheumatology*. 2023;62(6):2076-2082. doi:10.1093/rheumatology/keac560
14. Mathew S, Peat G, Parry E, Sokhal BS, Yu D. Applying sequence analysis to uncover 'real-world' clinical pathways from routinely collected data: a systematic review. *J Clin Epidemiol*. 2024;166:111226. doi:10.1016/j.jclinepi.2023.111226
15. Schwanitz K. The transition to adulthood and pathways out of the parental home: A cross-national analysis. *Adv Life Course Res*. 2017;32:21-34. doi:10.1016/j.alcr.2017.03.001
16. Lorentzen T, Bäckman O, Ilmakunnas I, Kauppinen T. Pathways to Adulthood: Sequences in the School-to-Work Transition in Finland, Norway and Sweden. *Soc Indic Res*. 2019;141(3):1285-1305. doi:10.1007/s11205-018-1877-4
17. Zhou Y. Work trajectories and status attainment process: a study using sequence analysis. *The Journal of Chinese Sociology*. 2023;10(1):1. doi:10.1186/s40711-022-00180-3
18. Brodeur S, Vanasse A, Courteau J, et al. Antipsychotic utilization trajectories three years after initiating or reinitiating treatment of schizophrenia: A state sequence analysis approach. *Acta Psychiatr Scand*. 2022;145(5):469-480. doi:10.1111/acps.13411
19. Vanasse A, Courteau J, Courteau M, et al. Multidimensional analysis of adult patients' care trajectories before a first diagnosis of schizophrenia. *Schizophrenia*. 2022;8(1):52. doi:10.1038/s41537-022-00256-6
20. Vanasse A, Courteau J, Courteau M, et al. Healthcare utilization after a first hospitalization for COPD: a new approach of State Sequence Analysis based on the "6W" multidimensional model of care trajectories. *BMC Health Serv Res*. 2020;20(1). doi:10.1186/S12913-020-5030-0
21. Henri S, Herrera R, Vanasse A, Forget A, Blais L. Trajectories of care in patients with chronic obstructive pulmonary disease: A sequence analysis. *Canadian Journal of Respiratory, Critical Care, and Sleep Medicine*. 2022;6(4):237-247. doi:10.1080/24745332.2021.1978907
22. Le Meur N, Vigneau C, Lefort M, et al. Categorical state sequence analysis and regression tree to identify determinants of care trajectory in chronic disease: Example of end-stage renal disease. *Stat Methods Med Res*. 2019;28(6):1731-1740. doi:10.1177/0962280218774811
23. Abbott A, Tsay A. Sequence Analysis and Optimal Matching Methods in Sociology. *Sociol Methods Res*. 2000;29(1):3-33. doi:10.1177/0049124100029001001
24. Liao TF, Bolano D, Brzinsky-Fay C, et al. Sequence analysis: Its past, present, and future. *Soc Sci Res*. 2022;107:102772. doi:10.1016/j.ssresearch.2022.102772
25. National Institute for Health and Care Excellence. Multiple long-term conditions: multimorbidity register. Accessed December 11, 2024. https://www.nice.org.uk/indicators/ind205-multiple-long-term-conditions-multimorbidity-register
26. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
27. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Med Care*. 1998;36(1):8-27. doi:10.1097/00005650-199801000-00004

28. Wray NP, Hollingsworth JC, Petersen NJ, Ashton CM. Case-Mix Adjustment Using Administrative Databases: A Paradigm to Guide Future Research. *Medical Care Research and Review*. 1997;54(3):326-356. doi:10.1177/107755879705400306
29. Ministry of Housing C& LG. The English Indices of Deprivation 2019 (IoD2019) . Accessed July 12, 2024. https://assets.publishing.service.gov.uk/media/5d8e26f6ed915d5570c6cc55/IoD2019_Statistical_Release.pdf
30. Hill JC, Kang S, Benedetto E, et al. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ Open*. 2016;6(8):e012331. doi:10.1136/bmjopen-2016-012331
31. Ritschard G, Liao TF, Struffolino E. Strategies for Multidomain Sequence Analysis in Social Research. *Sociol Methodol*. 2023;53(2):288-322. doi:10.1177/00811750231163833
32. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40-49. doi:10.1002/mpr.329
33. Meisingset I, Vasseljen O, Vøllestad NK, et al. Novel approach towards musculoskeletal phenotypes. *European Journal of Pain*. 2020;24(5):921-932. doi:10.1002/ejp.1541
34. Lynch M, Peat G, Jordan K, Yu D, Wilkie R. Where does it hurt? Small area estimates and inequality in the prevalence of chronic pain. *European Journal of Pain*. 2023;27(10):1177-1186. doi:10.1002/ejp.2148

Table 1: Patients' baseline characteristics.

Variables (n=1875)	
Gender, n (%)	
Female	1233 (65.76)
Male	642 (34.24)
Age, Mean (SD)	57.74 (15.50)
Age-group, n (%)	
18-34 years	157 (8.37)
35-44 years	234 (12.48)
45-54 years	362 (19.31)
55-64 years	430 (22.93)
65-74 years	407 (21.71)
75-84 years	241 (12.85)
85+ years	44 (2.35)
BMI, Mean (SD)	29.18 (6.91)
BMI, n (%)	
Underweight (<18.5)	26 (1.39)
Normal (18.5-24.9)	399 (21.28)
Overweight (25-29.9)	562 (29.97)
Obese (≥30)	608 (32.43)
Missing	280 (14.93)
Comorbidity count, n (%)	
0	829 (44.21)
1	577 (30.77)
2	329 (17.55)
3+	140 (7.47)
Index of Multiple Deprivation, n (%)	
Quintile 1 (most deprived)	530 (28.27)
Quintile 2	383 (20.43)
Quintile 3	398 (21.23)
Quintile 4	320 (17.07)
Quintile 5 (least deprived)	244 (13.01)
Ethnicity, n (%)	
White	1788 (95.36)
Asian	31 (1.65)
Mixed	11 (0.59)
Black	28 (1.49)
Other	17 (0.91)

Table 2. Multinomial logistic regression model for association between patients' characteristics and different clusters.

	Clusters of similar care sequences (Reference cluster is Low consultation and healthcare use)			
	Increasing consultation and analgesia	High consultation and healthcare use	Low consultation but high analgesia	Low consultation but moderate healthcare use
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Gender				
Male	1	1	1	1
Female	1.51 (0.97, 2.35)	**2.55 (1.69, 3.88)**	**1.79 (1.30, 2.46)**	**1.87 (1.40, 2.51)**
Age-group				
18-34 years	1	1	1	1
35-44 years	0.62 (0.23, 1.69)	1.24 (0.43, 3.57)	2.98 (0.96, 9.21)	1.41 (0.75, 2.62)
45-54 years	1.71 (0.76, 3.84)	**2.82 (1.10, 7.21)**	**5.73 (1.97, 16.66)**	1.80 (0.99, 3.24)
55-64 years	1.16 (0.50, 2.69)	**3.50 (1.39, 8.82)**	**7.04 (2.45, 20.21)**	**1.87 (1.04, 3.33)**
65-74 years	1.96 (0.85, 4.52)	**5.31 (2.09, 13.49)**	**11.66 (4.06, 33.53)**	**2.79 (1.55, 5.01)**
75-84 years	**2.73 (1.10, 6.77)**	**9.05 (3.41, 24.03)**	**18.99 (6.46, 55.84)**	**3.20 (1.68, 6.11)**
85+ years	1.69 (0.32, 8.89)	**8.42 (2.23, 31.80)**	**12.51 (3.26, 47.94)**	**4.31 (1.65, 11.26)**
BMI Underweight/Normal (<25)	1	1	1	1
Overweight (25-29.9)	0.98 (0.52, 1.84)	1.04 (0.59, 1.84)	1.18 (0.77, 1.80)	1.22 (0.84, 1.76)
Obese (≥30)	**2.03 (1.15, 3.59)**	**2.54 (1.52, 4.25)**	**1.79 (1.18. 2.71)**	**1.80 (1.25, 2.59)**
Comorbidity count				
0	1	1	1	1
1	**1.74 (1.07, 2.83)**	1.58 (0.97, 2.56)	**2.19 (1.51, 3.18)**	1.34 (0.99, 1.82)
2	**1.93 (1.07, 3.47)**	**4.38 (2.72, 7.05)**	**3.90 (2.60, 5.85)**	1.28 (0.86, 1.91)
3+	**3.49 (1.65, 7.38)**	**6.65 (3.60, 12.28)**	**5.55 (3.21, 9.61)**	**2.07 (1.19, 3.61)**
Index of Multiple Deprivation				
Quintile 1 (most deprived)	1.42 (0.71, 2.82)	**2.65 (1.34, 5.23)**	**2.09 (1.21, 3.62)**	1.22 (0.78, 1.90)
Quintile 2	1.24 (0.61, 2.52)	**2.33 (1.16, 4.65)**	**1.84 (1.05, 3.22)**	1.11 (0.70, 1.76)
Quintile 3	0.70 (0.33, 1.49)	0.78 (0.36, 1.66)	1.30 (0.75, 2.26)	0.85 (0.55, 1.35)
Quintile 4	0.86 (0.40, 1.83)	1.46 (0.70, 3.02)	1.34 (0.76, 2.38)	0.86 (0.53, 1.39)
Quintile 5 (least deprived)	1	1	1	1

Significant results are highlighted in bold. OR- Odds Ratio, BMI- Body Mass Index

Table 3. Longitudinal linear mixed model to assess association between clusters of similar care sequence and MSK-HQ score. Reference cluster is low consultation and healthcare use.

	MSK-HQ Score	
	Coefficients (95% CI)	P-value
Fixed effects		
Intercept	**26.19 (23.04, 29.33)**	**<0.001**
Cluster of similar care sequences		
Increasing consultation and analgesia	**-5.90 (-7.90, -3.89)**	**<0.001**
High consultation and healthcare use	**-7.26 (-9.01, -5.51)**	**<0.001**
Low consultation but high analgesia	**-5.79 (-7.24, -4.35)**	**<0.001**
Low consultation but moderate healthcare use	**-2.73 (-4.03, -1.43)**	**<0.001**
Time		
3-Months	**5.41 (4.78, 6.04)**	**<0.001**
6-Months	**6.42 (5.69, 7.16)**	**<0.001**
Interaction terms cluster of similar care sequence*time		
Increasing consultation and analgesia*3-Months	-2.05 (-4.12, 0.02)	0.053
Increasing consultation and analgesia*6-Months	**-2.82 (-3.30, -0.57)**	**0.024**
High consultation and healthcare use*3-Months	**-5.18 (-6.92, -3.43)**	**<0.001**
High consultation and healthcare use*6-Months	**-4.57 (-6.55, -2.58)**	**<0.001**
Low consultation but high analgesia*3-Months	**-2.88 (-4.28, -1.49)**	**<0.001**
Low consultation but high analgesia*6-Months	**-2.98 (-4.61, -1.35)**	**<0.001**
Low consultation but moderate healthcare use*3-Months	**-1.93 (-3.30, -0.57)**	**0.006**
Low consultation but moderate healthcare us*6-Months	-1.10 (-2.65, 0.45)	0.165
Random effects		
Intercept (SD)	7.39	
Time (SD)	2.76	

Model was controlled for gender, age, BMI, comorbidity count, and index of multiple deprivation. Significant results are highlighted in bold.

Table 4. Difference in MSK-HQ score from low consultation and healthcare use at baseline, 3-months, and 6-months.

	MSK- HQ score (Reference cluster is Low consultation and healthcare use)		
	Baseline	3-Months	6- Months
	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)
Increasing consultation and analgesia	**-5.90 (-7.91, -3.89)**	**-7.95 (-10.40, -5.49)**	**-8.72 (-11.61, -5.83)**
Low consultation and healthcare use	0	0	0
High consultation and healthcare use	**-7.26 (-9.01, -5.51)**	**-12.43 (-14.56, -10.31)**	**-11.82 (-14.24, -9.40)**
Low consultation but high analgesia	**-5.79 (-7.24, -4.35)**	**-8.68 (-10.39, -6.97)**	**-8.78 (-10.76, 6.79)**
Low consultation but moderate healthcare use	**-2.73 (-4.03, -1.43)**	**-4.66 (-6.28, -3.04)**	**-3.82 (-5.67, -1.98)**

Model was controlled for gender, age, BMI, comorbidity count, and index of multiple deprivation. Significant results are highlighted in bold.

Figure legends

Figure 1. State distribution plot of care sequence typology by domain (consultations, prescriptions, imaging, physiotherapy, and secondary care referrals)

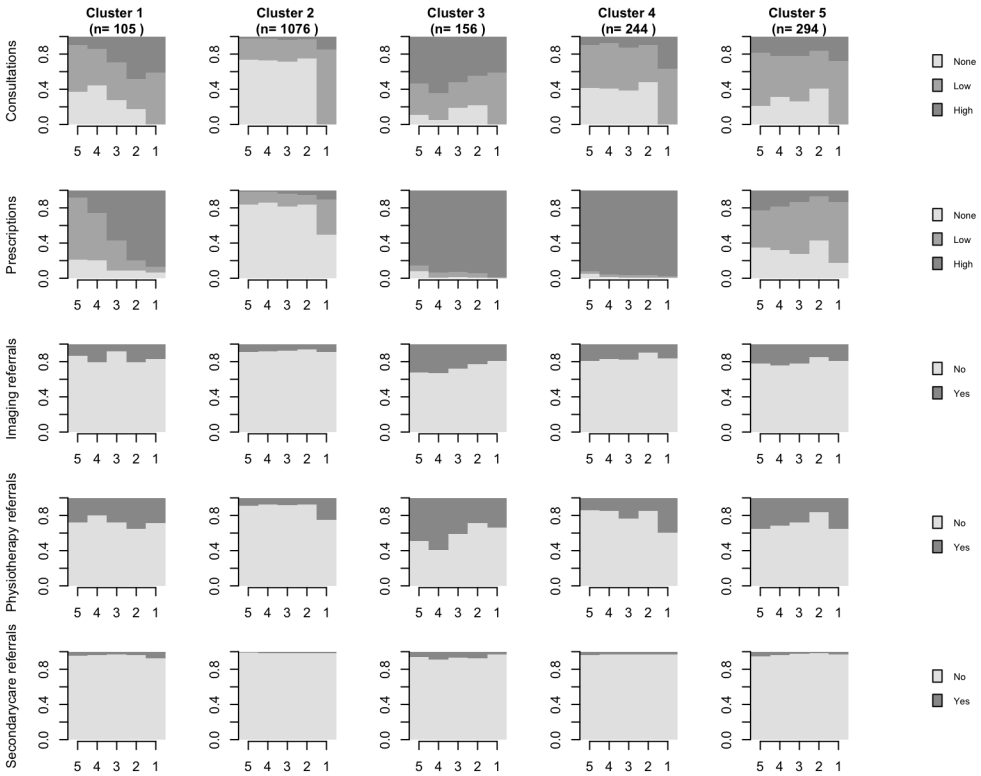

Figure 2. Predicted values of MSK-HQ score between the distinct clusters. Predicted values were controlled for gender, age, BMI, comorbidity count, and index of multiple deprivation.