
Article:
Smith, Chris orcid.org/0000-0002-0184-3590, Hewison, Jenny, West, Robert et al. (2 more authors) (2020) Understanding patterns of care for musculoskeletal patients using routinely collected National Health Service data from general practices in England. Health informatics journal. 2470–2484. ISSN 1460-4582

https://doi.org/10.1177/1460458220907431

Health Informatics Journal

Original Article

Health Informatics Journal
2020, Vol. 26(4) 2470–2484

DOI: 10.1177/1460458220907431
journals.sagepub.com/home/jhi
SAGE

Understanding patterns of care for musculoskeletal patients using routinely collected National Health Service data from general practices in England

Chris Smith
Leeds Institute of Health Sciences, University of Leeds

Jenny Hewison
Leeds Institute of Health Sciences, University of Leeds

Robert M West
Leeds Institute of Health Sciences, University of Leeds

Sarah R Kingsbury
Leeds Institute for Rheumatic and Musculoskeletal Medicine, University of Leeds

Philip G Conaghan
NIHR Leeds Musculoskeletal Biomedical Research Centre, University of Leeds

Abstract

Musculoskeletal conditions are extremely common and represent a costly and growing problem in the United Kingdom. Understanding patterns of care and how they vary between individual patients and patient groups is necessary for effective and efficient disease management. In this article, we present a novel approach to understanding patterns of care for musculoskeletal patients in which trajectories are constructed from clinical and administrative data that are routinely collected by clinicians and healthcare professionals. Our approach is applied to routinely collected National Health Service data for musculoskeletal patients who were registered to a set of general practices in England and highlights both known and previously unreported

Corresponding author:
Chris Smith, Leeds Institute of Health Sciences, University of Leeds, Level 10, Worsley Building, Clarendon Way, Leeds LS2 9NL, UK.
Email: c.j.smith@leeds.ac.uk

variations in the prescribing of opioid analgesics by gender and presence of pre-existing depression. We conclude that the application of our approach to routinely collected National Health Service data can extend the dimensions over which patterns of care can be understood for musculoskeletal patients and for patients with other long-term conditions.

Keywords
electronic health records, general practice, musculoskeletal disease, patterns of care

Introduction

Musculoskeletal conditions are extremely common and represent a costly and growing problem in the United Kingdom.[1] A total of 14.9 million people (29%) in England are estimated to live with a musculoskeletal condition.[1] In 2013–2014, musculoskeletal conditions accounted for the third largest area of National Health Service (NHS) programme spending at £4.7 billion.[2] It was estimated that the treatment of osteoarthritis and rheumatoid arthritis would account for £10.2 billion in direct costs to the NHS and wider healthcare system in 2018.[1]

Musculoskeletal disease is subject to gradual onset with symptoms increasing in frequency and severity over time. Risk factors such as pre-existing co-morbidities must be managed over the life course to reduce the risk of developing musculoskeletal disease and to manage disease progression.[1] Management is commonly undertaken in primary care,[1] and treatments include physical activity and pain management. Understanding patterns of care over time for musculoskeletal patients, such as prescriptions issued for chronic pain, and how they vary for individuals and groups would inform more effective and efficient disease management.

Understanding of patterns of care over time for musculoskeletal patients and patients with other long-term conditions has been previously limited by the time and cost constraints associated with project-specific data collection. However, administrative and clinical data are now routinely collected by clinicians and healthcare professionals[3] to inform patient care. With an appropriate ethical and legal basis, and robust governance arrangements in place, routinely collected data can offer a cost-effective source of observational data that can supplement or potentially replace project-specific data collection for clinical research.[4–6]

In this article, we present a novel approach to understanding patterns of care in which trajectories are constructed from clinical and administrative data that are routinely collected by clinicians and healthcare professionals. This approach was developed as part of a study to investigate factors that affect progression of musculoskeletal disease and is applied to routinely collected NHS data for musculoskeletal patients who were registered to set of general practices in England. Our results highlight both known and previously unreported variations in prescribing of opioid analgesics by gender and the presence of pre-existing depression for musculoskeletal patients. We conclude that the application of our approach to routinely collected NHS data can extend the dimensions over which patterns of care can be understood for musculoskeletal patients and for patients with other long-term conditions.

Material and methods

Ethical approval

Approval for the study was obtained from the School of Medicine Research Ethics Committee (SoMREC) at the University of Leeds (reference: SoMREC/13/079), and the Research Project Committee at ResearchOne (project number: 201428378A).

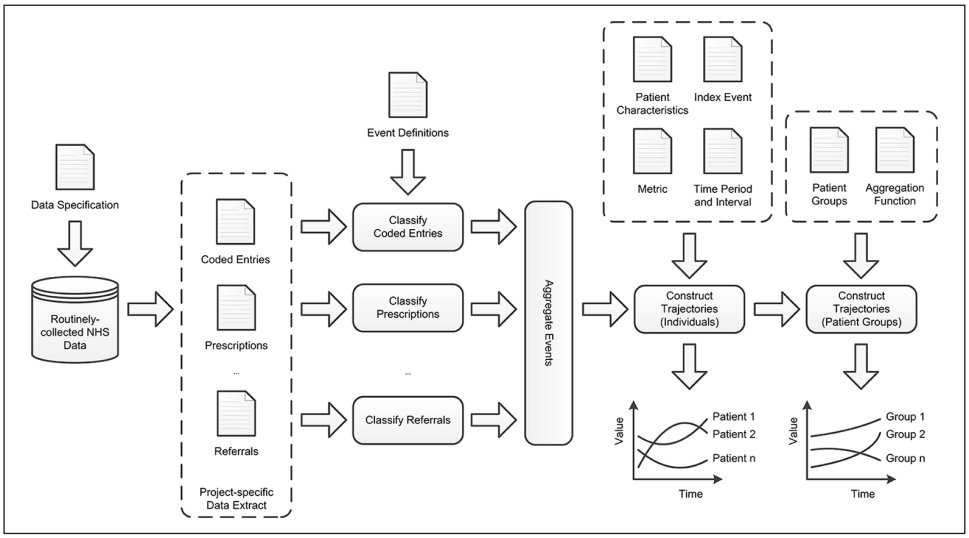

Figure 1. Workflow diagram illustrating the process by which trajectories are constructed for individual patients and patient groups from routinely collected NHS data.

Data

Routinely collected NHS data for the study were obtained from ResearchOne.[7] ResearchOne is a research database controlled by The Phoenix Partnership (TPP) that contains de-identified clinical and administrative data for patients who (1) are registered to general practices that use the SystmOne clinical information system[8] and which have opted-in to ResearchOne at practice-level, and (2) have not opted-out of ResearchOne at patient-level. All general practices that had opted-in to ResearchOne at the time of data extraction were located in England. Inclusion and exclusion criteria for the patient population and the data entries obtained for these patients were determined by the research team, which included a senior musculoskeletal clinician (P.C.).

Patient population. Patients were included who (1) were aged between 40 and 75, and (2) had their first record of a clinical code relating to joint pain between 1 April 1999 and 31 March 2014. Age criteria were determined by the ages over which patients are most likely to present with symptoms of musculoskeletal disease. The date 31 March 2014 represented the end of the last full financial year on commencement of the work and 1 April 1999 was chosen to provide up to 15 years of follow-up per patient from 31 March 2014. A total of 152,437 patients were referenced in the data obtained from ResearchOne.

Data entries. Selected clinical and administrative data entries relevant to the characterisation of musculoskeletal patients and their patterns of care for musculoskeletal disease were obtained for the period between 1 April 1999 and 31 March 2014. Clinical data entries included (1) coded diagnoses/observations, (2) prescriptions, (3) repeat prescriptions and (4) referrals. Administrative data entries included (1) practice registrations, (2) service interactions and (3) demographic data. Supplemental Material - Additional File 1 provides the definitions used to select relevant diagnoses/observations and prescriptions. Standardised sets of clinical codes defined within the Quality Outcomes Framework (QOF)[9] were used to define co-morbidities (see Figure 1).

Table 1. Example representation of events.

ID	Timestamp	Depression	Joint pain	. . .	Opioid
. . .	. . .	. . .	. . .	. . .	. . .
1	2000-09-13	1	0	. . .	1
1	2000-10-15	0	1	. . .	0
1	2000-11-15	0	1	. . .	1
1	2000-11-16	0	0	. . .	1
. . .	. . .	. . .	. . .	. . .	. . .

Table 2. Example representation of events following time normalisation.

ID	Timestamp	Depression	Joint pain	. . .	Opioid
. . .	. . .	. . .	. . .	. . .	. . .
1	−2	1	0	. . .	1
1	**1**	**0**	**1**	. . .	**0**
1	2	0	1	. . .	1
1	2	0	0	. . .	1
. . .	. . .	. . .	. . .	. . .	. . .

Example index event is highlighted in bold.

Event classification

Events were extracted from data entries by applying classification functions to the values of specific attributes in these entries. Classification functions were defined for events that were relevant to the characterisation of patients and their patterns of care for musculoskeletal disease. Application of these functions provided a representation of events relevant to care that was decoupled from their (varied) manifestations in the routinely collected NHS data[10–13] and which was homogeneous within and between patients. In addition to events contained within data entries, an event was explicitly included for each patient on 31 March 2014 to represent the date up to which data were received from ResearchOne.

Boolean outputs from classification functions were represented in a matrix. Columns were indexed by event and rows which were indexed by a unique project-specific patient identifier and the timestamp of the data entry (see Table 1). Occurrence of an event for a patient at a timestamp was represented with a 1 (True) value in the relevant matrix cell. Columns and rows containing only 0 (False) values were removed to reduce matrix dimensionality.

Time normalisation

Interaction with health services is not synchronised between patients.[14,15] Different patients are at different stages with respect to their care for a specific condition on a specific calendar date. To enable patterns of care over time to be meaningfully compared between different patients, timestamps (calendar dates) associated with the events of each patient were normalised with respect to an index event that was common to all patients. First recorded joint pain event (see Supplemental Material - Additional File 1) was chosen as the index event for this study as it was determined to represent a logical indication of the onset of musculoskeletal disease.

Table 3. Example representation for individual patients.

Patient characteristics					Prescribing days (opioid analgesics)					
Gender	Age	Depression	Backward	Forward	−4	. . .	−1	1	. . .	4
M	45	1	16	20	1	. . .	1	3	. . .	1
F	41	0	8	24	0	. . .	2	1	. . .	4
M	57	0	6	4	1	. . .	3	4	. . .	2
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
F	65	1	22	26	0	. . .	0	1	. . .	2

Patient characteristics are included along with values for an example metric – prescribing days (opioid analgesics) – at each time interval.

Normalisation replaced the timestamps of data entries with the number of days between the timestamp and the timestamp of the index event (see Table 2). Normalised time was represented in days due to the granularity of the timestamps associated with data entries. More coarse-grained representations of normalised time, such as months and quarters, were then derived from days. Events occurring before and after the index event were associated with negative and positive (normalised) timestamps, respectively.

Index events could not be determined for 5992 (3%) patients. ResearchOne determined that references to these patients had been included based on the fulfilment of inclusion criteria by data entries captured outside general practice. Required data entries had not been supplied for these patients, and they were omitted from any further consideration in the study.

Patient characterisation

To enable patterns of care to be compared within and between specific patient groups, patients were characterised by age, gender and the presence of 19 specific co-morbidities at the index event. Age was determined from the normalised time of a birth event and expressed in approximate years (360 days). Gender was straightforwardly determined from demographic data. Presence of co-morbidities was determined from occurrence of a relevant diagnosis/observation event (see Supplemental Material - Additional File 1) at any time in a period of 360 days prior to the index event. Dynamic (i.e. time-varying) clinical factors, such as the presence of co-morbidities, must be operationalised for analysis based on appropriate clinical and temporal constraints. Variation in how these factors are operationalised affects comparability between studies and requires careful consideration.

Two additional characteristics were also included for each patient that represented the (normalised) times up to which data entries were available before and after the index event (respectively) for that patient. We refer to these characteristics as *backward support* and *forward support*, respectively. Values for these characteristics were determined from the normalised time associated with the index event, birth event, death event (if applicable) and data extraction event.

Trajectories

Patterns of care were modelled as trajectories. Metrics were defined over patient events to provide a measure of a relevant dimension of care. Values were derived for these metrics at time intervals before and after the index event to form a trajectory. Changes in values between intervals were interpreted as changes in care received by the patient or patient group. Number of days comprising a time interval was varied to enable patterns of care to be explored at different time granularities, such as months and years.

Table 4. Example representation for patient groups.

Patient group		Mean prescribing days (opioid analgesics)					
Gender	Depression	−4	. . .	−1	1	. . .	4
M	0	0	. . .	1	1	. . .	2
M	1	1	. . .	2	3	. . .	5
F	0	1	. . .	2	1	. . .	2
F	1	2	. . .	3	5	. . .	6

Values of patient characteristics used to define the group are included along with values for an example metric – mean prescribing days (opioid analgesics) – at each time interval.

Individual patients. To understand patterns of care between individual patients, trajectories were initially constructed for individual patients (see Table 3). Metrics were defined for these patients based on the number of days on which six different classes of medication that are commonly used to treat musculoskeletal disease (see Supplemental Material - Additional File 1) were prescribed within a time interval. Prescriptions were chosen as the focus of the metrics for this study as they are generally subject to less variable recording practices in general practice than other dimensions of health and healthcare (e.g. referrals). All prescriptions of these medications were considered. We did not attempt to determine the specific indications for which medications were prescribed.

Values for the metrics were derived for four time intervals composed of different numbers of days: 30 ($\approxeq 1$ month), 90 ($\approxeq 1$ quarter), 180 ($\approxeq 0.5$ year) and 360 ($\approxeq 1$ year). By generating trajectories for different time intervals, we were able to compare the patterns of care observed at different intervals and to focus analysis on the time interval whose patterns were most amenable to clinical interpretation. For each time interval, values were defined over a time period of 360 days ($\approxeq 1$ year) before and 1800 days ($\approxeq 5$ years) after the index event. Patients required at least 360 days of backward support to be considered. Trajectories included values for intervals within this pre-index period. For each patient, a total of 24 (4 intervals * 6 metrics) different trajectories were constructed.

Patient groups. To understand patterns of care between patient groups, trajectories of individual patients with specific characteristics were used to construct trajectories for patient groups (see Table 4). Gender and presence of pre-existing depression have been previously shown to have an effect on patterns of care for musculoskeletal disease.[16] Therefore, analysis focused on patient groups defined by these characteristics. Trajectories were constructed for these groups for the same set of time intervals and time periods as individual patients. Metric values for each group at each time interval were determined from the application of a specific aggregation function (mean) to the metric values of all patients in the group at that time interval. Group members without sufficient forward and backward support for a trajectory defined over a specific time period were omitted to prevent distortion of the group metric values at earlier and later time intervals.

Results and discussion

Patient characteristics

Table 5 summarises the characteristics that were derived for the musculoskeletal patients who were included in the study. A total of 85,575 (60.5%) of musculoskeletal patients were female. Joint Pain was first recorded between the age of 50 and 75 for over 90 per cent of male and female patients.

Coronary heart disease was present in 3.6 per cent of male patients compared with 1.6 per cent of female patients. Depression was present in 3.5 per cent of female patients compared with 1.9 per cent of male patients. Hypothyroidism was present in 1.9 per cent of female patients compared with 0.5 per cent of male patients. A total of 31.5 per cent of male patients had ever smoked compared to 22 per cent of female patients. Over 75 per cent of male and female patients had no co-morbidities present when Joint Pain was first recorded. Male and female patients had a median backward support of 34 intervals (3060 days $\approxeq$ 8.5 years) and 33 intervals (2970 days $\approxeq$ 8.25 years), respectively, and a median forward support of 26 intervals (2340 days $\approxeq$ 6.5 years) and 27 intervals (2430 days $\approxeq$ 6.75 years), respectively.

Trajectories (Individual patients)

Table 6 summarises the trajectories that were constructed for the musculoskeletal patients who were included in the study. Variation in patterns of care between patients is represented straightforwardly by the number of unique trajectories, where a unique trajectory is a unique set of values for a metric over the defined set of time intervals and time period. Higher variation between patients for a given time period, time interval and metric is represented by a higher number of unique trajectories.

Number of patients (N) with sufficient forward support decreases with the post-index period. N decreases by 30 per cent between 360 days ($\approxeq$ 1 year) and 1800 days ($\approxeq$ 5 years). In addition, 5.6 per cent of the 141,346 patients had insufficient forward support from which to construct a trajectory with a post-index period of 360 days ($\approxeq$ 1 year). Such reductions in the patient population due to the time periods for which data are required present a significant challenge for robust analysis of long-term conditions, and illustrate the importance of the inclusion/exclusion criteria used to select patients and their data items in retaining the feasibility of certain analyses.

Number of unique trajectories increases with post-index period for any given interval size across all metrics. Larger post-index periods increase the number of values that comprise a trajectory for any given interval size, and therefore increase the dimensions of the value space from which a trajectory can be drawn. Number of unique trajectories decreases with increases in interval sizes for any given post-index period across all metrics. Larger interval sizes reduce the number of values that comprise a trajectory for any given post-index period, and therefore reduce the dimensions of the value space from which a trajectory can be drawn. Such variations in the number of unique trajectories illustrate the importance of time periods and intervals in determining the space of trajectories that constructed for subsequent analysis.

Trajectories based on the prescribing days of non-steroidal anti-inflammatory drugs (NSA), opioid analgesics (OPI) and non-opioid analgesics and compound analgesic preparations (NOP) exhibit the largest number of unique trajectories over all time periods and intervals. Prescriptions for these medications are issued on a greater number of days per time interval across all patients than prescriptions for other medications. This increases the upper bound on the space of metric values for a specific time interval. Such variation in the number of unique trajectories illustrates the importance of metrics in determining the space of trajectories that are constructed for subsequent analysis.

Trajectories (Patient groups)

Trajectories were constructed for groups of musculoskeletal patients defined by gender and the presence of pre-existing depression from the characteristics and trajectories that were previously constructed for individual patients. Trajectories were constructed for metrics based on the mean

Table 5. Summary of characteristics for musculoskeletal patients who were registered to a set of general practices in England.

	Male	Female
Number of patients	55,771	85,575
Age at index event, n (%)		
40–49	4748 (8.5)	6982 (8.2)
50–59	18,053 (32.4)	28,324 (33.1)
60–69	18,485 (33.1)	25,980 (30.4)
70+	14,485 (26.0)	24,289 (28.4)
Presence of condition at index event, n (%)		
Asthma	1229 (2.2)	2702 (3.2)
Atrial fibrillation	565 (1.0)	470 (0.5)
Cancer	579 (1.0)	747 (0.9)
Coronary heart disease	1877 (3.4)	1339 (1.6)
Chronic kidney disease	895 (1.6)	1600 (1.9)
Chronic obstructive pulmonary disease	1166 (2.1)	1311 (1.5)
Dementia	75 (0.1)	148 (0.2)
Depression	1049 (1.9)	2915 (3.4)
Diabetes	2592 (4.6)	2813 (3.3)
Epilepsy	129 (0.2)	207 (0.2)
Heart failure	209 (0.4)	197 (0.2)
Hypertension	4844 (8.7)	7713 (9)
Hypothyroidism	268 (0.5)	1667 (1.9)
Learning disability	26 (0.0)	20 (0.0)
Osteoporosis	87 (0.2)	892 (1.0)
Peripheral arterial disease	192 (0.3)	128 (0.1)
Psychosis, schizophrenia, bipolar affective	96 (0.2)	189 (0.2)
Stroke	206 (0.4)	219 (0.3)
Stroke (TIA)	195 (0.3)	194 (0.2)
Number of conditions present at index event, n (%)		
0 condition	42,832 (76.8)	65,111 (76.1)
1 condition	10,198 (18.3)	16,350 (19.1)
2 conditions	2239 (4.0)	3364 (3.9)
3 or more conditions	502 (0.9)	750 (0.9)
Smoking status at index event, n (%)		
Ever smoked	17,543 (31.5)	18,797 (22)
Support from index event, median intervals (IQR)		
Backward support	34 (21–46)	33 (21–45)
Forward support	26 (14–39)	27 (15–39)

TIA: transient ischemic attack; IQR: interquartile range.

prescribing days of the six different classes of medication for which the trajectories of individual patients were constructed. We focus our results on mean prescribing days for opioid analgesics due to questions that have been raised about the efficacy of opioid analgesics in treating long-term pain.[17,18] Trajectories were constructed for 1800 days ($\approx$ 5 years) after the index event, and 360 days ($\approx$ 1 year) before the index event to assist interpretation with respect to our definition of pre-existing depression (see 'Patient characterisation' section).

Table 6. Summary of trajectories for musculoskeletal patients for different post-index time periods, time intervals and metrics.

Post-index period (days)	Interval (days)	Patients N	Number of unique trajectories					
			NSA	RUB	OPI	COR	NOP	DSR
360	30	133,313	22,314	8251	17,396	709	27,219	1093
	90		10,889	4669	10,553	273	15,297	765
	180		3175	1980	4359	118	4802	489
	360		385	352	746	42	682	191
720	30	124,230	30,230	11,429	22,220	1244	33,460	1493
	90		21,377	8067	17,106	587	26,483	1276
	180		12,362	5181	11,711	293	16,966	996
	360		2815	1868	3991	115	4287	559
1080	30	113,853	34,282	13,316	24,774	1690	35,685	1652
	90		26,384	9998	20,095	915	30,422	1492
	180		19,246	7344	16,006	504	24,628	1317
	360		9123	4037	9388	207	12,632	1000
1440	30	103,115	36,680	14,621	26,134	1997	36,089	1671
	90		29,299	11,282	21,769	1159	31,479	1531
	180		22,894	8726	18,021	696	27,206	1393
	360		14,524	5612	12,903	326	19,338	1193
1800	30	92,115	37,345	15,335	26,466	2255	35,394	1662
	90		30,688	12,092	22,388	1409	31,293	1522
	180		24,911	24,911	19,021	889	26,647	1409
	360		17,486	6675	14,585	464	22,014	1234

Metrics relate to prescriptions of the following medications – NSA: non-steroidal anti-inflammatory drugs; RUB: rubefacients, topical NSAIDs, capsaicin and poultices; OPI: opioid analgesics; COR: corticosteroids; NOP: non-opioid analgesics and compound analgesic preparations; DSR: drugs that suppress the rheumatic disease process.

Permutation tests[19,20] were used to determine statistical significance of the observed variations between patient groups. Observed variations between patient groups were compared with 9999 random assignments of group labels to patients. Individual trajectories remained unaltered. Note that multiple significance tests have been employed. A significance level of 5 per cent or 1 per cent might be intended and following any appropriate adjustment for multiple tests, such as Bonferroni,[21] the reported results of the permutation tests will remain highly significant.

Variation by gender. Figure 2 (top left) illustrates the variation in trajectories by gender: *there is strong evidence that female musculoskeletal patients receive more opioid analgesics than male musculoskeletal patients* ($p < 0.0001$). Previous work has shown that more women than men are prescribed analgesia,[22,23] corroborating our result. Our trajectories also illustrate a general increase in prescribing of opioid analgesics over normalised time. General increases in prescribing for opioid analgesics have been previously demonstrated in calendar time.[24] Our work demonstrates such increases with respect to normalised time and therefore life course.

Variation by presence of pre-existing depression. Figure 2 (top right) illustrates the variation in trajectories by the presence of pre-existing depression: *there is strong evidence that musculoskeletal patients with pre-existing depression receive more opioid analgesics than musculoskeletal patients*

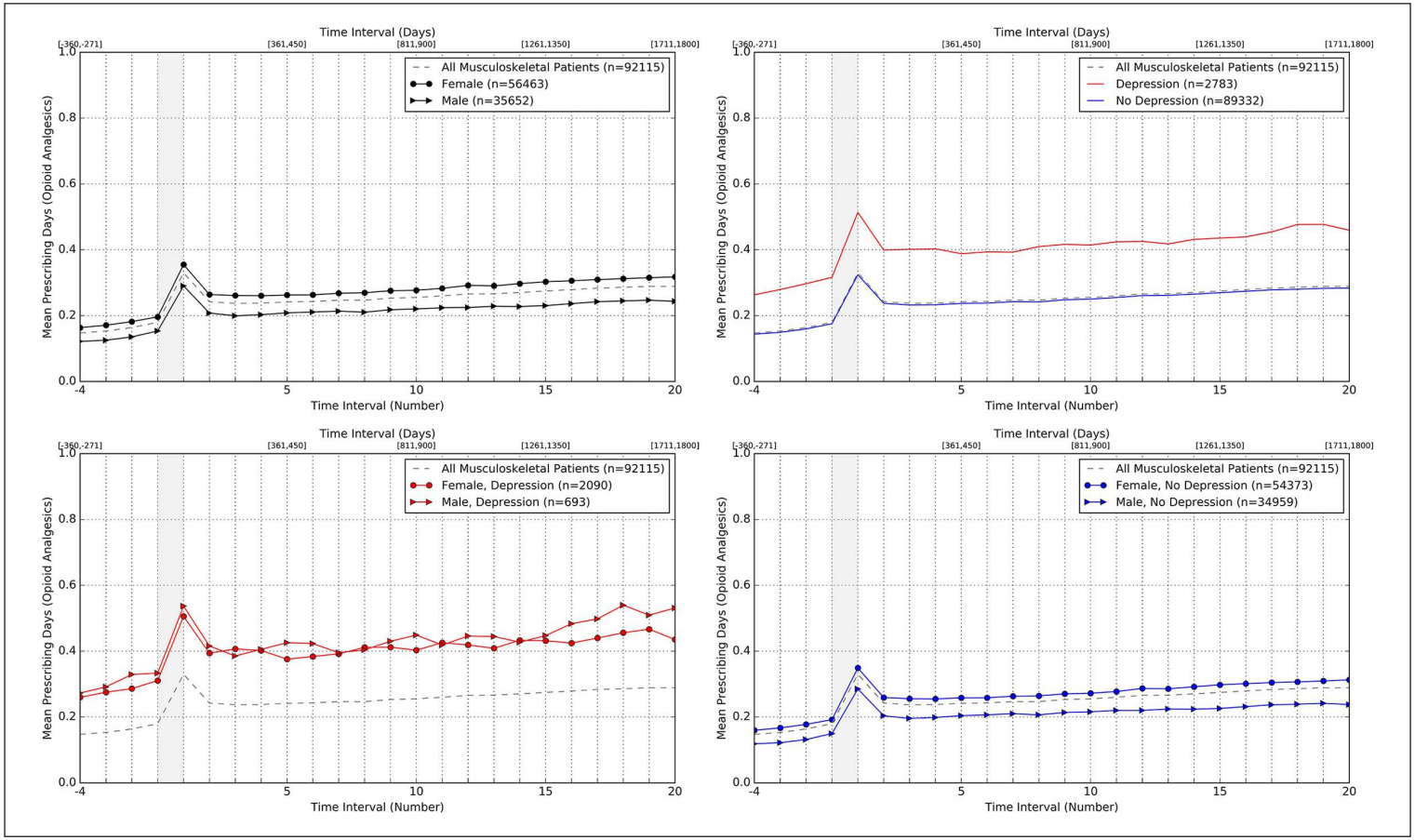

Figure 2. Trajectories based on mean prescribing days (opioid analgesics) for groups of musculoskeletal patients characterised by gender (top left); presence of pre-existing depression (top right); gender and presence of pre-existing depression (bottom left); and gender and no pre-existing depression (bottom right).

Table 7. Definition and description of changes in prescribing state based on the value, v, of any two time intervals, t_i and t_j where $t_i > t_j$.

t_i	t_j	Description
$v = 0$	$v = 0$	Continuing non-prescribed
$v = 0$	$v > 0$	Newly prescribed
$v > 0$	$v = 0$	Newly non-prescribed
$v > 0$	$v > 0$	Continuing prescribed

without pre-existing depression ($p < 0.0001$). Previous work has shown that depression is a risk factor for pain and the prescribing of opioid analgesics[16,25–27] corroborating our result.

Variation by gender and absence of pre-existing depression. Figure 2 (bottom right) illustrates the variation in trajectories by gender for musculoskeletal patients without pre-existing depression: *there is strong evidence that female musculoskeletal patients without pre-existing depression receive more opioid analgesics than male musculoskeletal patients without pre-existing depression* ($p < 0.0001$). This is consistent with the effect of gender alone, which is illustrated in Figure 2 (top left).

Variation by gender and presence of pre-existing depression. Figure 2 (bottom left) illustrates the variation in trajectories by gender for patients with pre-existing depression: *there is no strong evidence of a difference between genders in the receipt of opioid analgesics for musculoskeletal patients with pre-existing depression* ($p = 0.2965$). Pre-existing depression appears to change the effect of gender on the prescribing of opioid analgesics. Differential effects of this nature have not been widely reported to date and are worthy of further investigation.

Trajectories constructed for groups of musculoskeletal patients illustrate a general increase in prescribing of opioid analgesics over normalised time. Clinicians concerned about the overall rise in prescribing of opioid analgesics understandably focus on long-term users. However, trajectories based on mean prescribing days cannot show whether increases are attributable to (1) an increase in the amount of prescriptions to patients already prescribed opioid analgesics, and/or (2) an increase in the proportion of patients who receive a prescription for opioid analgesics.

To differentiate these effects, and to demonstrate the ability to define inter-interval metrics, a metric was defined to represent change in the 'state' of opioid analgesic prescribing for each patient at each time interval. Table 7 provides the definition of this metric based on the value, v, of any two time intervals, t_i and t_j where $t_i > t_j$. Trajectories were constructed for patients based on this metric for 1800 days ($\approxeq 5$ years) after the index event and 360 days ($\approxeq 1$ year) before the index event. Trajectories were then constructed for patient groups defined by gender and the presence of pre-existing depression to illustrate the proportion of patients in the group in a particular state at each time interval.

Figure 3 illustrates the proportion of patients with a particular change in prescribing state over time for each of the four patient groups. The proportion of musculoskeletal patients who are newly prescribed increases at the index event across all groups. The proportion of patients who are newly *non*-prescribed then increases in the subsequent time interval. This indicates that patients within each group are often prescribed opioid analgesics for a short time period ($\approxeq 90$ days) in response to the index event. However, the proportion of musculoskeletal patients who are prescribed increases over time for all patient groups, and the proportion of male musculoskeletal patients with pre-existing depression who are prescribed is consistently higher than any of the other patient groups over the time period.

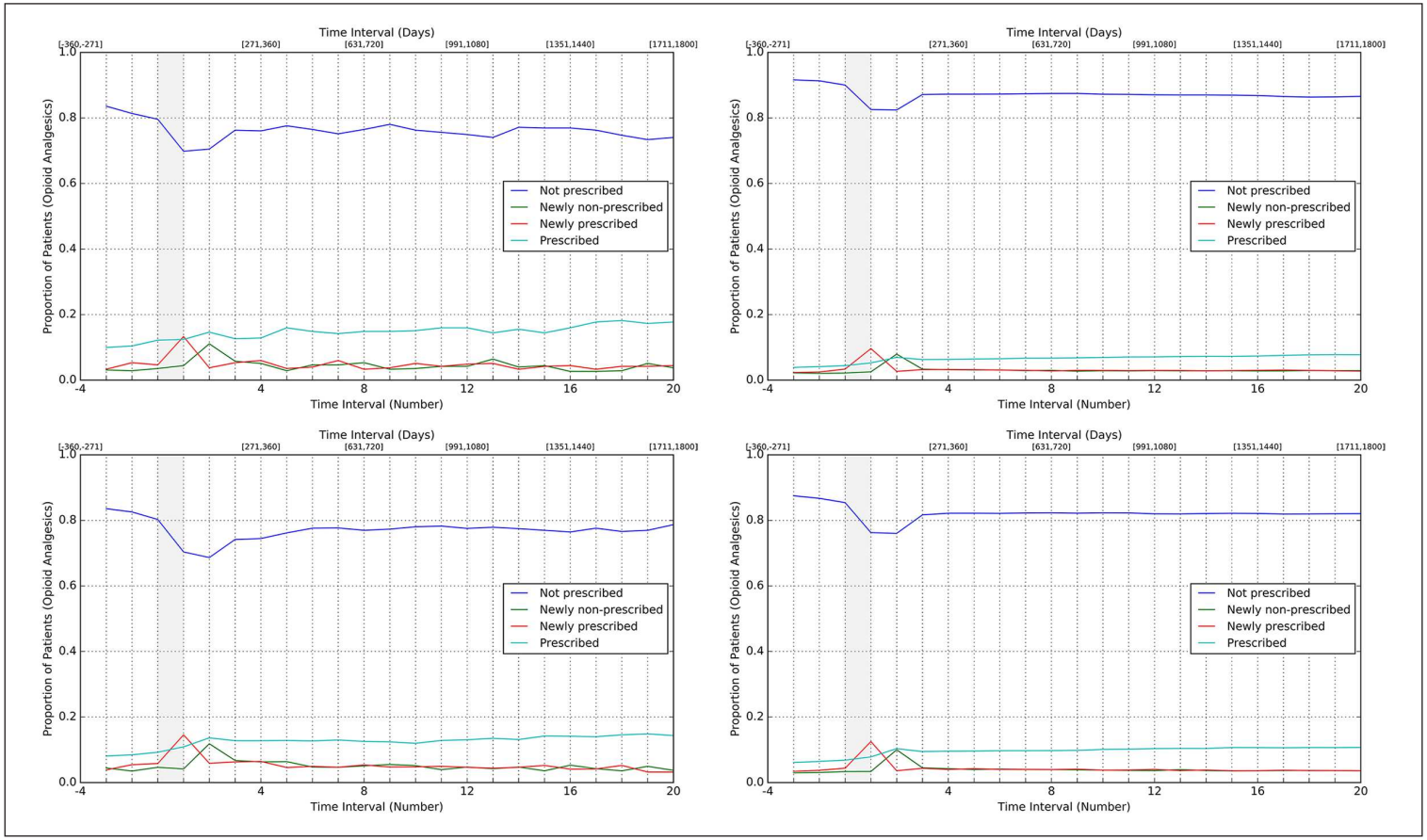

Figure 3. Trajectories representing proportion of musculoskeletal patients within each group who had a particular change in prescribing state for opioid analgesics between subsequent time intervals. Groups shown are male with depression (top left); male with no depression (top right); female with depression (bottom left); female with no depression (bottom right).

This implies that the increases illustrated for each group are not solely due to increases in the frequency of prescriptions for group members who were previously prescribed opioid analgesics, but due to increases in the number of group members who receive a prescription for opioid analgesics.

Conclusion

We have presented a novel approach to understanding patterns of care in which trajectories are constructed from clinical and administrative data that are routinely collected by clinicians and healthcare professionals. The approach was applied to data for musculoskeletal patients who were registered to a set of general practices in England and highlighted both known and previously unreported variations in prescribing of opioid analgesics by gender and presence of depression.

Strengths

Any dimensions of health and healthcare that are routinely collected by clinicians and healthcare professionals in electronic format can be used to characterise patients and to understand their patterns of care using our approach. Analysts can iteratively explore trajectories for patients with specific characteristics, or can apply methods such as latent class analysis[28] to determine those characteristics associated with specific patterns of care.

Our approach is independent of the specific data model(s) to which the routinely collected data conform – classification functions can be defined to classify events from any data model and subsequent processing steps are then analogous. The approach is also independent of the long-term condition for which patterns of care are to be constructed – subject to the availability of the required data, events, metrics and characteristics can be defined and trajectories can be constructed to understand patterns of care for any long-term condition.

Normalisation of time enables patterns of care over time to be decoupled from calendar time such that patterns are not simply artefacts of a particular period of (calendar) time. While cross-sectional study designs enable differences between individual patients and patient groups to be discovered over time, our approach retains the same set of patients over time and enables time intervals over which metric values are determined to be varied.

Validity of our approach is demonstrated through correspondence of our results with known variations in prescribing of opioid analgesics from clinical literature. Our results also contribute previously unreported variations between specific patient groups that are worthy of further investigation.

Limitations

Construction of trajectories is subject to significant computational overhead, which increases with the number of patients and time intervals. Data quality issues are common in routinely collected NHS data (e.g. missingness and inconsistency) and must be considered when defining events, metrics, characteristics and time intervals. For instance, such issues present a significant challenge in determining causal relationships between events, such as symptoms and the subsequent prescription of medications. In addition, both clinical and technical inputs are required to ensure the robust definition of representative index events, metrics and characteristics for the specific long-term condition to be studied.

Characteristics of the underlying data and the operationalisation of clinical definitions from the data introduce significant complexity to the comparison of results between studies – motivating a

rigorous approach to study documentation and provenance. Any interpretation of trajectories must also consider the inherent limitations of observational data, which are captured outside of experimental conditions,[29,30] such as inherent biases. We conclude that our approach can extend the dimensions over which patterns of care can be understood for musculoskeletal patients and for patients with other long-term conditions.

Acknowledgements

The authors are thankful to ResearchOne which provided data used for this study. The authors are also thankful to Professor Allan House who contributed to the clinical interpretation of the trajectories.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: C.S. is Director of PrivacyForge Limited. All other authors declare that they have no conflicting interests.

Ethical approval

Approval for the study was obtained from the School of Medicine Research Ethics Committee (SoMREC) at the University of Leeds (reference: SoMREC/13/079), and the Research Project Committee at ResearchOne (project number: 201428378A).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by (1) The Leeds Teaching Hospitals NHS Trust through the Applied Health Cooperative at the Leeds Institute of Health Sciences; (2) National Institute for Health Research (NIHR) through the Leeds Musculoskeletal Biomedical Research Centre; and (3) Yorkshire and Humber Commissioning Support Research Capability Funding (RCF) (reference: RCF-2014-005). This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or Department of Health.

ORCID iD

Chris Smith https://orcid.org/0000-0002-0184-3590

Supplemental material

Supplemental material for this article is available online.

References

1. Arthritis UK. State of Musculoskeletal Health 2018: arthritis and other musculoskeletal conditions in numbers. Report, 2019.
2. NHS England. 2013/2014 CCG programme budgeting benchmark tool, https://www.england.nhs.uk/wp-content/uploads/2015/06/13-14-ccg-prog-bug-benchmarking-tool.xlsm (2014, accessed 13 January 2020).
3. Häyrinen K, Saranto K and Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform* 2008; 77(5): 291–304.
4. Jensen PB, Jensen LJ and Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012; 13(6): 395–405.
5. Powell J and Buchan I. Electronic health records should support clinical research. *J Med Internet Res* 2005; 7(1): e4.

6. Ramakrishnan N, Hanauer D and Keller B. Mining electronic health records. *Computer* 2010; 43(10): 77–81.
7. The Phoenix Partnership (TPP). ResearchOne, http://www.researchone.org (2020, accessed 13 January 2020).
8. The Phoenix Partnership (TPP). SystmOne, https://www.tpp-uk.com/products/systmone (2020, accessed 13 January 2020).
9. NHS Digital. Quality and outcomes framework, https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/quality-outcomes-framework-qof (2020, accessed 13 January 2020).
10. de Lusignan S. The barriers to clinical coding in general practice: a literature review. *Med Inform Internet Med* 2005; 30(2): 89–97.
11. Tai TW, Anandarajah S, Dhoul N, et al. Variation in clinical coding lists in UK general practice: a barrier to consistent data entry? *Inform Prim Care* 2007; 15(3): 143–150.
12. Andrews JE, Richesson RL and Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc* 2007; 14(4): 497–506.
13. Jordan K, Porcheret M and Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam Pract* 2004; 21(4): 396–412.
14. Batal I, Sacchi L, Bellazzi R, et al. A temporal abstraction framework for classifying clinical temporal data. *AMIA Annu Symp Proc* 2009; 2009: 29–33.
15. Krstajić M, Bertini E and Keim DA. Cloudlines: compact display of event episodes in multiple time-series. *IEEE Trans Vis Comput Graph* 2011; 17(12): 2432–2439.
16. Gambassi G. Pain and depression: the egg and the chicken story revisited. *Arch Gerontol Geriatr* 2009; 49(Suppl. 1): 103–112.
17. British Medical Association. Chronic pain: supporting safer prescribing of analgesics. Report, 2017, https://www.bma.org.uk/collective-voice/policy-and-research/public-and-population-health/analgesics-use
18. Public Health England. Dependence and withdrawal associated with some prescribed medicines: an evidence review. Report, 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/829777/PHE_PMR_report.pdf
19. Fisher R. *The design of experiments*. New York: Hafner, 1935.
20. Pitman EJG. Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika* 1938; 29(3–4): 322–335.
21. Dunn OJ. Multiple comparisons among means. *J Am Stat Assoc* 1961; 56(293): 52–64.
22. Foy R, Leaman B, McCrorie C, et al. Prescribed opioids in primary care: cross-sectional and longitudinal analyses of influence of patient and practice characteristics. *BMJ Open* 2016; 6(5): e010276.
23. Hansen CA, Inacio MC, Pratt NL, et al. Chronic use of opioids before and after total knee arthroplasty: a retrospective cohort study. *J Arthroplasty* 2017; 32(3): 811–817.
24. Thielke SM, Simoni-Wastila L, Edlund MJ, et al. Age and sex trends in long-term opioid use in two large American health systems between 2000 and 2005. *Pain Med* 2010; 11(2): 248–256.
25. Rosemann T, Gensichen J, Sauer N, et al. The impact of concomitant depression on quality of life and health service utilisation in patients with osteoarthritis. *Rheumatol Int* 2007; 27(9): 859–863.
26. Rosemann T, Laux G, Szecsenyi J, et al. Pain and osteoarthritis in primary care: factors associated with pain perception in a sample of 1,021 patients. *Pain Med* 2008; 9(7): 903–910.
27. Serdarevic M, Striley CW and Cottler LB. Sex differences in prescription opioid use. *Curr Opin Psychiatry* 2017; 30(4): 238–246.
28. West R, Hill K, Hewison J, et al. Psychological disorders after stroke are an important influence on functional outcomes: a prospective cohort study. *Stroke* 2010; 41(8): 1723–1727.
29. Weiskopf NG, Hripcsak G, Swaminathan S, et al. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform* 2013; 46(5): 830–836.
30. Weiskopf NG, Rusanov A and Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA Annu Symp Proc* 2013; 2013: 1472–1477.