BMJ Open

Patterns of care for people presenting to Australian general practice with musculoskeletal complaints based upon routinely collected data: Protocol for an observational cohort study using the Population Level Analysis and Reporting (POLAR) database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055528
Article Type:	Protocol
Date Submitted by the Author:	16-Jul-2021
Complete List of Authors:	Haas, Romi; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine Busija, Ljoudmila; Monash University, Department of Epidemiology and Preventive Medicine Gorelik, Alexandra; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine O'Connor, Denise; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine Pearce, Christopher; Outcome Health Mazza, Danielle; Monash University, Department of General Practice Buchbinder, Rachelle; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine
Keywords:	PRIMARY CARE, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, Back pain < ORTHOPAEDIC & TRAUMA SURGERY, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Knee < ORTHOPAEDIC & TRAUMA SURGERY, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Patterns of care for people presenting to Australian general practice with musculoskeletal complaints based upon routinely collected data: Protocol for an observational cohort study using the Population Level Analysis and Reporting (POLAR) database

Romi Haas (romi.haas@monash.edu) [1,2*], Ljoudmila Busija (lucy.busija@monash.edu) [2], Alexandra Gorelik (Alexandra.gorelik@monash.edu) [1,2], Denise O'Connor (denise.oconnor@monash.edu) [1,2], Christopher Pearce (drchrispearce@mac.com) [3], Danielle Mazza (danielle.mazza@monash.edu) [4], Rachelle Buchbinder (rachelle.buchbinder@monash.edu) [1,2]

[1] Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Health, Malvern, Victoria, Australia.
[2] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia.
[3] Outcome Health, Blackburn, Victoria, Australia.
[4] Department of General Practice, Monash University, Notting Hill, Victoria, Australia.
[*] Corresponding author

Correspondence

Romi Haas
Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Institute
154 Wattletree Rd, Malvern
Victoria 3144
Australia
Email: romi.haas@monash.edu
Telephone: +61 3 9508 3499

Keywords
Protocol
Primary care
Musculoskeletal disorders
Low back pain
Neck pain
Shoulder pain
Knee pain

ABSTRACT

Introduction

General practice is integral to the Australian healthcare system. Outcome Health's POpulation Level Analysis and Reporting (POLAR) database uses de-identified electronic health records to analyse general practice data in Australia. Previous studies using routinely collected health data for research have not consistently reported the codes and algorithms used to describe the population, exposures, interventions and outcomes in sufficient detail to allow replication. This paper reports a study protocol investigating patterns of care for people

presenting with musculoskeletal conditions to general practice in Victoria, Australia. Its focus is on the systematic approach used to classify and select eligible records from the POLAR database to facilitate replication. This will be useful for other researchers using routinely collected health data for research.

Methods and analysis

This is a retrospective cohort study. Patient-related data will be obtained through electronic health records from a subset of general practices across three primary health networks (PHN) in South Eastern Victoria. Data for patients with a low back, neck, shoulder and/or knee condition and who received at least one GP face-to-face consultation between 1/01/2014 and 31/12/2018 will be included. Data quality checks will be conducted to exclude patients with poor data recording and/or non-continuous follow-up. Relational data files with eligible and valid records will be merged to select the study cohort and the GP care received (consultations, imaging requests, prescriptions and referrals) between diagnosis and 31/12/2018. Number and characteristics of patients and GPs, and number, type and timing of imaging requests, prescriptions for pain relief and referrals to other health providers will be investigated.

Ethics and dissemination

Ethics approval was obtained from the Cabrini and Monash University Human Research Ethics Committees (Reference Numbers 02-21-01-19 and 16975 respectively). Study findings will be reported to Outcome Health, participating PHNs, disseminated in academic journals and presented in conferences.

ARTICLE SUMMARY

Strengths and limitations of this study

- This is the first study to our knowledge to report the codes and algorithms used to classify, select and merge eligible records from the POLAR database into a patient-centred database to facilitate analysis of general practice patterns of care.
- The systematic approach used in this study can be adapted by other researchers using routinely collected health data for research purposes.
- This study will extend previous research that has assessed the representativeness of POLAR data to GP care across the wider Australian population.
- These data are likely to underestimate actual allied health visits as some of these do not require a GP referral in Australia; some prescriptions for pain relief are available without a prescription so these data will also be underestimated.
- It is possible not all patterns of care for the study cohort will be directly attributable to a musculoskeletal condition as reasons for GP consultations, referrals, and prescriptions are not mandated by the source Electronic Medical Records (EMRs).

INTRODUCTION

General practice plays an essential role in providing primary health care to the population. In Australia 86% of the population visits a general practitioner (GP) multiple times a year[1], and nearly 20% of these consultations are for a musculoskeletal condition[2]. These conditions account for 23% of the years lived with disability in Australia[3] and are also a major cause of disability worldwide[4]. Until 2016, the BEACH (Bettering the Evaluation and Care of Health) program provided the most comprehensive data on clinical activities of Australian general practice[5]. The program identified a number of activities that represent low-value care for people with musculoskeletal conditions including an over-reliance on imaging, prescription of opioids, and unnecessary referrals to specialist care[6 7]. However, in-depth exploration of these activities within the BEACH program is limited by its cross-sectional design, and these data are no longer being collected.

Technological advancements have facilitated the extraction of de-identified patient information from general practice clinical information systems. The advantage of these datasets for research purposes are that they are longitudinal and can therefore be used to establish sequences of events at the patient level and to examine changes in GP management over time. Both the Medicine Insight[8] and the Population Level Analysis and Reporting (POLAR) databases[9] are examples of longitudinal general practice datasets within Australia. Unlike POLAR, the Medicine Insight program does not currently include referrals provided by GPs to other healthcare providers[8]. These data may provide important insights into how well GPs are playing their role as gatekeepers of the Australian healthcare system.

While using routinely collected data for research purposes offers considerable opportunities to improve health care, there are several challenges to be overcome. Differences in patient information management and data extraction tools result in variability in both the information captured and ways in which this information is coded. In particular, the way in which text values (diagnoses, examination findings, test results and medications) are transformed to codes can be a source of variation within and between studies. Previous studies have highlighted how code selection affects the reported prevalence and precision of results[10]. Studies conducted using routinely collected health data should therefore be reported with sufficient detail and clarity to allow replication. However, a systematic evaluation of a random sample of 124 publications using routinely collected health data has demonstrated inadequate reporting of the methods used[11]. For example, in 44 studies where definitions of codes or classification algorithms were deemed necessary to describe the population, exposures or interventions and outcomes, only nine (20.5%) reported all three items adequately. The REporting of studies Conducted using Observational Routinely collected Data (RECORD) guidelines, published in 2015, were developed to assist in this process and to ensure that readers can assess the internal and external validity of the findings of these studies[12].

The POLAR database draws data from every consultation occurring for millions of patients in approximately 30% of general practices across South-Eastern Victoria[13], an area that comprises more than half of Victoria's population[14]. Inclusion is based on practice consent so this volume is increasing exponentially as more practices consent to add their data and as more consultations occur over time. Unlike in other countries, coding is not embedded in the clinical process and needs to be conducted specifically for research purposes. Data are provided to research users in a relational database that organises data into files that can be merged based on common data fields. Identifying and selecting relevant records and merging separate files

into a patient-centred database for analysis is a complex task that could potentially yield variable results depending on the methods used.

Previous studies have used the POLAR database to investigate patterns of antimicrobial prescribing for children[15], to examine characteristics of patients presenting to an after-hours clinic[16], to estimate GP recording of cardiovascular risk factors[17], and to describe characteristics of pathology test ordering in general practice[18]. However, these studies have not reported the methods used to classify and select eligible records or the processes used to merge data files into a patient-centred database for analysis.

This manuscript presents a protocol for a study investigating patterns of GP care for people with a low back, neck, shoulder and/or knee condition in Victoria, Australia. It describes the methods used to classify and select eligible records from the POLAR database and how relational data files will be merged into a patient-centred database. This systematic approach will guide future research by enabling researchers interested in using routinely collected health data, and the POLAR database in particular, to answer other clinically relevant questions about general practice care. Study findings will advance existing knowledge about GP care for people with these musculoskeletal conditions and whether it conforms to best evidence-based practice. Differences in care across different musculoskeletal complaints may also inform tailored interventions to improve care and ultimately reduce the burden of disease associated with these musculoskeletal complaints.

Objectives

The aim of this study will be to examine GP patterns of care for people with low back, neck, shoulder and knee conditions. Specific objectives will be to:

1. Describe and compare the management (number, type and timing of imaging tests and procedure requests, prescriptions for pain relief, and referrals to other health providers) provided by GPs to people with low back, neck, shoulder and knee conditions
2. Describe the prevalence of comorbidities among specific musculoskeletal diagnoses within this cohort
3. Examine the association between management types and patient- and practice-related variables
4. Examine the longitudinal changes in GP management for these conditions between 2014 and 2018 inclusive

METHODS

Study design

A retrospective cohort study using general practice health records from Victoria, Australia.

Data source

This study will use data from Outcome Health's POLAR database[9]. The database structure is based on eight relational files, each containing de-identified practice, provider, and/or patient codes (Figure 1). These common fields allow merging of the data files so that databases can be configured for specific research purposes. Data is extracted from two different clinical information systems, covering ninety percent of included general practices. All data is extracted using the Hummingbird data extraction tool[9].

Setting

The POLAR database contains de-identified patient-related data from all electronic medical records of consenting general practices within the PHNs of Eastern Melbourne, South Eastern Melbourne and Gippsland within Victoria, Australia. Our study will include data collected over five calendar years from 1 January 2014 until 31 December 2018 relating to all patients with an eligible musculoskeletal condition and who received at least one face-to-face GP consultation. Follow up will be from the time of the initial recorded diagnosis to 31st December 2018. Data analyses will be completed by the end of 2021.

Participants

The study cohort will include people diagnosed during 2014 to 2018 inclusive with a low back, neck, shoulder and/or knee condition, limited to age 45 years and over except for low back which will be limited to age 18 years and over. The differing age restrictions were chosen because the prevalence of most musculoskeletal conditions increases markedly after the age of 45 except for low back pain which increases after the age of 18[19]. Eligibility criteria are presented (Table 1). We excluded traumatic diagnoses and conditions typically primarily managed by a specialist (e.g., inflammatory and autoimmune rheumatic diseases). Patients with an eligible diagnosis and age will also have received at least one GP face-to-face consultation during the study dates. The musculoskeletal diagnosis will not have to occur during a GP consultation since it is an eligible diagnosis that could result from consultation with other health care providers.

Variables

Preparatory work to classify and select eligible records has been completed as part of the protocol process. In circumstances where Outcome Health has previously coded data (e.g., diagnosis records), we used this coding to select eligible records that fitted our inclusion criteria. In circumstances where there was no coding (e.g., imaging tests), we coded the data into categories and then selected eligible records. Outcome Health's approach to coding used clinical natural language processing to automatically code structured narrative text within the electronic medical record following by a manual process for quality checking and correction[20]. For example, this allowed the free text items 'back pain', 'low back pain', and 'lumbar pain' to all sit under the same diagnostic code. Where possible, coding was conducted using a standardised classification system. For example, diagnoses are coded using SNOMED CT-AU terminology[21] and prescriptions are coded according to the Anatomical Therapeutic Chemical (ATC) classification system[22]. In cases where there is no standardised classification system available (e.g., providers and referrals), Outcome Health used a similar process to code these variables into relevant categories (e.g., type of health care provider). Clinical natural language processing conducted by Outcome Health has previously demonstrated accurate coding of over 95% of the narrative text to SNOMED CT-AU terms in a sample of approximately 57,000 diagnosis records[20]. Our approaches to coding and/or selecting eligible records for each variable are described in detail below.

Provider records

Healthcare providers other than a GP may be nested within a general practice. To limit all diagnoses, consultations, referrals, and prescriptions to those made only by GPs we used coding within the provider type field conducted by Outcome Health. This is coded by Outcome Health

according to the professional background of the healthcare provider delivering the service (e.g., GP, nurse).

Diagnoses records

All SNOMED CT-AU diagnosis-related terms used during 2014-2018 were searched by two study authors (RH and RB) to select eligible low back, neck, shoulder and knee conditions. We included all patients with an eligible musculoskeletal diagnosis during 2014-2018 regardless of whether they had a prior musculoskeletal diagnosis. Included SNOMED diagnosis terms are presented (Table 2). Sacral conditions were included as part of low back conditions. The following SNOMED terms were excluded as these conditions were deemed to be indicative of traumatic injury or conditions that are not managed primarily by GPs: fracture (except lumbar and tibial plateau fractures), dislocation, synovectomies/synovitis, and cauda equina syndrome. Knee ligamentous and meniscal tears were included as these are likely due to degeneration in the 45 years and over age group[23]. Lesions were excluded as these could involve a wound, ulcer or tumour and are not musculoskeletal conditions. General musculoskeletal terms such as sprain or osteoarthritis (where the site was not specified) were also excluded as these could not be attributed to a specific body region. We included relevant surgical or procedural musculoskeletal terms as GPs are involved in referral and follow-up for these conditions.

Using experienced clinicians, Outcome Health has further categorised SNOMED diagnoses into overarching groups and utilised key chronic disease groups as a qualifier[9]. For example, free text such as 'low back pain' or 'angina' could be qualified as a chronic disease if present for six months or more. We used these chronic disease groups to identify eligible comorbid diagnoses for our study cohort as follows: chronic cardiovascular disease, chronic obstructive pulmonary disease, chronic musculoskeletal conditions, cancer, opioid addiction, dementia, diabetes, depression/anxiety, and obesity. Obesity was identified using SNOMED terms as it was not coded as a chronic disease category in the POLAR database. We included previous chronic musculoskeletal conditions so that these could be investigated as a potential predictor of different management patterns.

Activity records

Activity records are coded in POLAR according to the type of consultation provided (e.g., telehealth, visit, telephone). Each time a note is recorded in the narrative section it is coded by the EMR and this is extracted by POLAR. We used this coding to select eligible patients who had at least one 'Activity type' relating to a face-to-face consultation (i.e., encounter, surgery or visit) during 2014-2018 inclusive. Telehealth and telephone consultations were also included for follow-up consultations only.

Referral records

Referral records are coded in POLAR according to discipline (e.g., neurosurgeon, physiotherapist, endocrinologist). We used this coding to select eligible referral groups considered relevant to a person with low back, neck, shoulder or knee conditions. The following referral groups were included: orthopaedics and neurosurgery (surgical specialists); sports medicine, rheumatology, rehabilitation medicine, neurology, and pain management (non-surgical specialists); and physiotherapy, osteopathy, massage therapy, exercise physiology, chiropractor, and psychology (allied health providers).

Prescription records

Medications are coded in POLAR according to the Anatomic and Therapeutic Classifications (ATC) system[22]. We included medications deemed by the study authors to be commonly prescribed for pain relief to people with musculoskeletal conditions. Medications within the following categories were included: simple analgesics; non-steroidal anti-inflammatories (NSAIDs); chondroitin and/or glucosamine; topical products for joint and muscular pain; opioids; neuromodulators and any relevant combinations. We included neuromodulators such as gabapentin and pregabalin because these are being increasingly used for the management of musculoskeletal conditions such as nonspecific low back pain or sciatica despite evidence of a lack of effectiveness and a higher risk of adverse events[24]. Opioid analgesics were further categorised into (i) weak single ingredient opioid analgesics (e.g. codeine), defined as <50 morphine milligram equivalents (MME) per day; (ii) strong single ingredient opioid analgesics (e.g. tapentadol, oxycodone, morphine), defined as 150 MME per day; and (iii) combination opioid analgesics[25]. Medicines in the combination opioid category were categorised based on the strongest medicine present, either as a weak combination opioid or as a strong combination opioid.

To ensure we included all potentially eligible medication names, we searched by both ATC category and by medication name from the prescription file during 2014-2018. The medication names we included are presented in Table 3. We included oral, topical and injectable preparations of medications. We excluded the following prescriptions: aspirin, decongestants (e.g., pseudoephedrine), antihistamines (e.g., doxylamine), opioid cough suppressants (e.g., dextromethorphan), and expectorants (e.g., guaifenesin). These were excluded on the basis that they were likely to have been prescribed for another condition (e.g., aspirin for secondary prevention of cardiovascular disease[26]).

Imaging records

The test data file within POLAR contains radiology and pathology tests requested by the GP. At the time of data extract, coding of the test data file had not been completed for specific imaging tests by Outcome Health and there were too many records to scan manually. We therefore exported all radiology test names during 2014-2018 inclusive and used an inductive coding process to select the following eligible imaging tests: plain radiographs, CT and MRI scans of the lumbar and cervical spine; plain radiographs, CT, MRI and ultrasounds of the knee; and plain radiographs, MRI scans and ultrasounds of the shoulder. We also included lumbar spine, knee, shoulder and cervical spine injections and shoulder hydrodilatation as eligible radiology procedures.

To code eligible imaging records, we first used the string match command in Stata to select all test names for each eligible anatomical region (i.e., low back, neck, shoulder and knee). Within each region, we then iteratively coded all imaging records into subgroups according to the type of imaging test (e.g., ultrasound). This process involved developing string match terms to identify each type of eligible radiology test or procedure within the sample, reviewing the uncoded test names (subgrouped as 'other') and manually coding additional terms until the remaining test names could not be classified into any further subgroups. We also developed string match terms to identify bilateral tests of the shoulder and knee. The initial string match terms used to code each body region and eligible imaging test or procedure are presented in Appendix 1.

During the coding process, there were numerous test names that did not definitively identify a type of imaging test (e.g., 'right knee'). We labelled these as 'unspecified'. We plan to classify these as plain radiographs in our analysis. This is because plain radiograph was deemed to be the default radiology modality in the Electronic Medical Record (EMR) software. The subgroups of imaging records inductively developed for each eligible body region are presented in Table 4. Our subgroup coding (excluding test names labelled as 'unspecified' and 'other') accounted for 96%, 95.8%, 95.2% and 96.6% of the identified low back (n=180,630), neck (n=192,844), shoulder (n=236,803) and knee (n=235,123) imaging test names respectively.

Test names indicating more than one imaging test were classified separately. We excluded imaging tests of soft tissues of the neck and test names indicating a combined neck image with the head, larynx, thyroid and/or abdomen (unless it specifically stated cervical spine) as we deemed these investigations were most likely not requested for a musculoskeletal condition. We also excluded test names with the following terms as these were not deemed to indicate an imaging test or procedure: 'report', 'findings', 'cancel', 'results', 'letter'.

Data access and cleaning

Outcome Health provided the research team with access to all POLAR database records since inception (1997). Data quality checks will be performed to label data as 'acceptable' for analysis using a similar process to that conducted by an established general practice database in the United Kingdom[27]. Duplicate data and records with empty or implausible birth dates (defined as greater than 115 years of age at time of diagnosis or dated after patient management) will be excluded from analyses. We will exclude practices without any activity data during 2014-2018. We will also examine the consistency of activity, test, prescription, and referral data for each practice in each eligible calendar year. If a gap in reporting from any practice is identified for one year or more, only data from the earliest date after which there was no gap will be included. For example, if a practice has activity data in 2014, 2017 and 2018, only data from 2017 onwards will be included. In addition, we will exclude activity records that represent more than one face-to-face consultation with a GP for the same patient on the same day. This is because an 'activity' occurs in POLAR anytime a patient record is accessed regardless of whether this was for clinical or administration purposes.

Approach to dataset creation

We will use a systematic process to systematically exclude ineligible records in order to merge data and select the study cohort (Figure 2). This process will require the merging of five relational data files (patient, practice, provider, activity and diagnosis) in a specific sequence to ensure all relevant records are retained. For example, we will not limit diagnosis records to 2014-2018 until after we have selected relevant comorbidities. A patient-centred database will be prepared to examine the number and type of GP consultations, imaging test and procedure requests, prescriptions for pain relief, and referrals to other health providers for our study cohort. Data that does not match our eligibility criteria (including data with missing fields) will be excluded during the merging process as unmatched records. Duplicate records, records with implausible dates or missing fields, and multiple records of the same type on a single day will also be removed and reported.

Analyses

All relevant data will be extracted from the POLAR SQL database and imported into Stata 15 (STATA Corp LP, College Station, TX, USA) for data management and analyses. The methods in this protocol are structured according to RECORD guidelines (Appendix 2)[12]. Full lists of codes used to define eligible variables are available from https://clinicalcodes.rss.mhs.man.ac.uk/medcodes/article/174/[28] where there is a recognised coding system.

Descriptive statistics will be used to summarise the study cohort including the number and type of eligible musculoskeletal conditions, patient demographics and comorbidities. These will be compared to national health survey data to assess the representativeness of the POLAR database to the wider Australian population. Eligible musculoskeletal conditions will be grouped according to body region.

Primary analysis will include analysis of each management type provided for each participant during the first year after their index diagnosis. A sensitivity analysis will be conducted including the entire follow-up period until 31st December 2018. For prescriptions, the primary analysis will include the entire follow-up period because repeated prescriptions over more than one year are anticipated. Descriptive statistics will also be used to summarise the number and type of GP all-cause consultations, imaging tests and procedures requested, prescriptions for pain relief, and referrals to other health providers for the study cohort. Results will be stratified by affected body region. Consultations will be categorised as face-to-face or telecommunication. Imaging requests will be categorised according to the type of imaging modality or procedure and body region (e.g., knee MRI). Bilateral knee and shoulder imaging requests will be counted as two imaging requests. Prescriptions will be categorised according to paracetamol, NSAIDs, glucosamine and/or chondroitin, opioids (weak single opioid, strong single opioid, weak combination opioid, and strong combination opioid) and neuromodulators. Referrals will be categorised according to surgical specialist, non-surgical specialist, and allied health. Patterns and timing of management (imaging requests, prescriptions and referrals) for people with eligible low back, neck, shoulder and knee conditions will be examined and compared between each year within the five-year study period and relative to time of diagnosis using trend analyses.

One of the limitations of the POLAR database is that it does not capture reasons for the clinical encounter or management types (imaging request, prescription or referral). To account for the subsequent uncertainty in attributing management types to a particular diagnosis for those with multiple musculoskeletal conditions, participants with eligible musculoskeletal diagnoses from multiple body regions will be analysed separately to those with eligible diagnoses in one body region. For participants with multiple eligible musculoskeletal diagnoses throughout the study period, the primary analysis will be conducted relative to the date of the first (index) eligible musculoskeletal diagnosis and a sensitivity analysis relative to the date of the most recent diagnosis will also be conducted. Imaging requests will be analysed relative to the date of the most recent musculoskeletal diagnosis for the same body region. For example, a shoulder ultrasound will be analysed

relative to the index date of an eligible shoulder diagnosis even if the same patient was diagnosed previously with an eligible knee condition.

The association between management types and patient- and practice-related characteristics will be examined using regression analysis. Predictors will include patient age, gender, body region(s) affected by eligible musculoskeletal conditions, socioeconomic status, remote or metropolitan location of GP practice, whether the patient lives within the Primary Health Network (PHN) of the practice, and time since diagnosis. Socioeconomic status will be defined by the Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD) using 2016 Census data[29].

Sequence analysis will be used to categorise sequences of management types of people with eligible musculoskeletal conditions into similar groups based on observed characteristics[30]. This will take into account both the time since diagnosis and sequence of each management type. We will use this to identify the most frequently used combinations and sequences of management and the patient- and practice-related variables that correlate with each management combination.

Sample size consideration

Sequence analysis will require the largest sample size of our planned analyses and will therefore form the basis of our sample size consideration. We plan to examine the following six management types: non-surgical referrals, surgical referrals, allied health referrals, opioid prescription, X-ray and/or ultrasound requests, and MRI and/or CT scan requests. This provides a total of 720 potential sequence combinations. Based on a recommended 20 to 30 subjects per subgroup[31], we estimate a sample size of between 14,400 and 21,600 will be required to differentiate between each sequence combination or pattern of care. Recent use of the POLAR database using data from approximately 200 general practices identified 20,514 active adult patients with type 2 diabetes before July 2016[32]. Our extract is based on 301 general practices from 2014 to 2018 and since the prevalence of diabetes is less than that of musculoskeletal conditions[33], we expect a sample size of more than 20,000.

Patient and public involvement

There will be no involvement of patients or the public in this study.

DISCUSSION

Explicitly reporting our systematic approach used to classify, select and merge eligible records from relational data files into a patient-centred database for analysis promotes transparency, reproducibility and completeness of the reporting of research conducted using routinely collected health data. The approach used to code eligible imaging tests from structured narrative text coded over 95% of the 845,400 cumulative imaging-related test and procedure records identified for low back, shoulder, knee and neck conditions during 2014-2018. Our code lists are available for all variables that have been previously coded by POLAR and those with a recognised coding system have been made available on the ClinicalCodes online repository. Although our coding process may only be applicable to systems that do not embed

coding in the clinical process, this approach can also be adapted to examine patterns of care over time for other conditions in general practice.

The main strength of this study is that it will facilitate an overview of the care provided by GPs to the same patient(s) over time and thereby enable temporal sequences to be examined. The POLAR database contains all patient-related activity within each practice making it representative of the included practices. Previous research has demonstrated comparable prevalence and age-gender distribution of people diagnosed with type 2 diabetes within the POLAR database to those within Australia[32]. This study will add to these findings by assessing the representativeness of people with musculoskeletal conditions within the POLAR database to the wider Australian population.

Constraints within the POLAR database may potentially limit the reliability of this study's findings although these are problems inherent in the use of any extracted data. Variability in workflows and recording behaviour introduces potential biases and the different clinical information systems used by the practices within POLAR may result in variability in the information entered. The objective of POLAR is to remove as much variability as possible by using and being transparent about the coding process. High accuracy of diagnostic coding by Outcome Health has been previously demonstrated[20]. In addition, it is possible not all patterns of care for the study cohort will be directly attributable to a musculoskeletal condition because reasons for GP consultations, referrals, and prescriptions are not mandated in the source EMRs. Prescriptions for some types of pain relief and referrals to allied health providers may also be underestimated by the POLAR database as these forms of management may be generated outside of general practice. These data are also likely to underestimate actual allied health visits and prescriptions for pain relief as some of these do not require a GP referral and are available over-the-counter without a prescription respectively in Australia.

ETHICS AND DISSEMINATION

Prior approval to conduct this study was obtained from the Cabrini Human Research Ethics Committee and Monash University Human Research Ethics Committee (Reference Numbers 02-21-01-19 and 16975 respectively). We did not obtain participant consent as all data was anonymised. Outcome Health holds a standing ethics approval for its collection and custodianship of the data from the Royal Australian College of General Practice. The study findings will be reported to Outcome Health, participating PHNs, disseminated in peer-reviewed academic journals and presented in national and international conferences.

FIGURES

Figure 1. Database structure
Figure 2. Approach to dataset creation

Table 1. Eligibility Criteria

Patient population			Patient management				
Diagnoses	**Provider**	**Patient**	**Practice**	**Activity**	**Referrals**	**Prescriptions**	**Imaging tests & procedures**
Low back Knee Shoulder Neck **Exclude:** Trauma Systemic inflammatory arthritis	Diagnosed by a general practitioner	Aged ≥18 years for low back conditions Aged ≥45 years for all other diagnoses	Patient activity 2014-2018	Face-to-face Telehealth	Surgical specialists Non-surgical specialists Allied health providers e.g. psychologist	Simple analgesics Anti-inflammatories Chondroitin/ Glucosamine Topical products Opioids Neuromodulators	Lumbar plain radiograph Lumbar CT Lumbar MRI Lumbar injection Knee plain radiograph Knee CT Knee MRI Knee ultrasound Knee injection Shoulder plain radiograph Shoulder ultrasound Shoulder MRI Shoulder injection Shoulder hydrodilatation Cervical plain radiograph Cervical CT Cervical MRI Cervical injection

Table 2: Included SNOMED terms

Low back diagnoses	Knee diagnoses	Shoulder diagnoses	Neck diagnoses
Arthritis of spine	Acute meniscal tear, medial	Acromioclavicular joint structure	Cervical arthritis
Arthropathy of spinal facet joint	Anterior knee pain	Adhesive capsulitis of shoulder	Cervical arthrodesis
Back problem	Arthritis of knee	Arthritis of acromioclavicular joint	Cervical disc disorder
Backache	Arthrodesis of knee	Arthrodesis of shoulder	Cervical kyphosis
Bone structure of coccyx	Arthroscopic lateral patellar release	Arthrography of shoulder	Cervical laminectomy
Bone structure of L5	Arthroscopic meniscectomy	Arthroscopic acromioplasty	Cervical myelopathy
Bone structure of sacrum	Arthroscopic procedure	Arthroscopic shoulder decompression	Cervical nerve root compression
Chondrectomy of spine	Arthroscopy of knee	Arthroscopy of shoulder	Cervical radiculitis
Chronic back pain	Arthroscopy of knee with lateral meniscectomy	Bursitis of shoulder	Cervical radiculopathy
Chronic lower back pain	Arthroscopy of knee with medial meniscectomy	Calcific tendinitis	Cervical rib
Compression fracture	Arthrotomy of knee	Calcific tendinitis of shoulder	Cervical spinal fusion by anterior technique
Compression fracture of vertebral column	Aspiration of knee joint	Capsulitis	Cervical spine degeneration
Compression of lumbar nerve root	Both knees	Contusion of shoulder region	Cervical spine structure
Correction of scoliosis	Bursitis of knee	Detachment of the glenoid labrum and/or capsule of the shoulder joint	Cervicogenic headache
Crush fracture of lumbar vertebra	Calcium pyrophosphate deposition disease	Entire tendon of supraspinatus muscle	Cervico-occipital neuralgia
CT of lumbar region	Chondrocalcinosis	Full thickness rotator cuff tear	Chronic neck pain
CT of lumbar spine	Chondromalacia of patella	Impingement syndrome of shoulder region	CT of cervical spine
CT of spine	Complete tear, knee, medial collateral ligament	Inflammation of rotator cuff tendon	CT of neck
Curvature of spine	Contusion of knee	Injury of glenoid labrum of shoulder joint	Degeneration of cervical intervertebral disc
Decompression laminectomy	Derangement of knee	Injury of shoulder region	Diffuse cervicobrachial syndrome
Decompression of lumbar spine	Disorder of patellofemoral joint	MRI of shoulder	Excision of cervical intervertebral disc
Degeneration of intervertebral disc	Finding of tear meniscus	Osteoarthritis of acromioclavicular joint	Injury of cervical spine
Degeneration of lumbar intervertebral disc	Fracture of tibial plateau	Osteoarthritis of shoulder	Kyphoscoliosis deformity of spine
Diagnostic radiography of coccyx	Haemarthrosis of knee	Painful arc syndrome	Kyphosis deformity of spine
Discitis	Inflammation of bursa of patella	Radiography of shoulder	Magnetic resonance imaging of neck
Discogenic pain			MRI of cervical spine

Disorder of joint of spine	Injury of anterior cruciate ligament	Repair of musculotendinous cuff of shoulder	Muscle spasm of cervical muscle of neck
Disorder of vertebra	Injury of knee	Repair of shoulder	Neck injury
Exploration of spine	Knee joint - varus deformity	Rotator cuff impingement syndrome	Neck pain
Facet joint pain	Knee joint effusion	Rotator cuff syndrome	Neck sprain
Fracture of body of vertebra	Knee joint valgus deformity	Rupture of tendon of biceps	Neck structure
Fracture of lumbar spine	Knee locking	Rupture of tendon of biceps, long head	Pain in cervical spine
Fracture of sacrum	Knee pain	Shoulder pain	Prolapsed cervical intervertebral disc
Fracture of vertebral column	Knee region structure	Shoulder reconstruction	Radiography of cervical spine
Injury of back	Knee stiff	Shoulder region structure	Spinal stenosis in cervical region
Injury of coccyx	Loose body in knee	Shoulder strain	Stiff neck
Intervertebral disc prolapse	MRI of knee	Shoulder tendinitis	Strain of neck muscle
L4/5 disc	Osteoarthritis of knee	Sprain of acromioclavicular ligament	Strain of tendon of neck
L5/S1 disc	Osteotomy of proximal tibia	Sprain of shoulder	Torticollis
Laminectomy	Osteotomy of tibia	Structure of left shoulder region	Whiplash injury to neck
Lordosis deformity of spine	Patellar instability	Structure of right shoulder region	
Low back pain	Patellar maltracking	Structure of rotator cuff including muscles and tendons	
Low back strain	Patellar tendonitis	Subacromial bursitis	
Lower back injury	Patellectomy	Subdeltoid bursitis	
Lower back structure	Patellofemoral osteoarthritis	Subluxation of acromioclavicular joint	
Lumbar	Patellofemoral stress syndrome	Subscapularis tendinitis	
Lumbar discectomy	Prepatellar bursitis	Supraspinatus tear	
Lumbar laminectomy	Problem knee	Supraspinatus tendinitis	
Lumbar microdiscectomy	Radiologic examination of knee	Total shoulder replacement	
Lumbar radiculopathy	Repair of anterior cruciate ligament of knee joint	US shoulder region	
Lumbar region back structure	Repair of knee collateral ligaments		
Lumbar spinal fusion	Repair of knee cruciate ligaments		
Lumbar sprain	Repair of meniscus		
Lumbosacral spine	Repair of patellar tendon		
Lumbosacral spondylosis	Replacement of total knee joint		
Lumbosacral spondylosis without myelopathy	Rupture of anterior cruciate		

Lumbosacral strain Lumbosacral radiculopathy Magnetic resonance imaging of spine Manipulation of spine MRI of lumbar spine Nerve root compression syndrome Nerve root disorder Operative procedure on spinal structure Osteoarthritis of lumbar spine Pain in lumbar spine Pain in the coccyx Prolapsed lumbar intervertebral disc Radiography of spine Sacral back pain Sacroiliac arthrodesis Sacroiliac joint inflamed Sacroiliac joint pain Scoliosis deformity of spine Scoliosis of lumbar spine Spasm of back muscles Spinal arthritis deformans Spinal arthrodesis Spinal claudication Spinal injury Spinal stenosis Spinal stenosis of lumbar region Spondylitis Spondylolisthesis	ligament Rupture of cruciate ligaments Rupture of medial collateral ligament of knee Rupture of posterior cruciate ligament Sprain of knee Sprain of lateral collateral ligament of knee Sprain of medial collateral ligament of knee Stabilisation of patellofemoral joint Strain of knee Strain of patellar tendon Strain of tendon of medial thigh muscle Structure of left knee Structure of prepatellar bursa Structure of right knee Subluxation of patellofemoral joint Suprapatellar bursitis Swollen knee Synovial cyst of knee Synovial cyst of popliteal space Tear of lateral meniscus of knee Tear of medial meniscus of knee Tear of meniscus of knee Total knee replacement Total replacement of left knee joint Total replacement of right knee joint Traumatic rupture of patellar		

Spondylolisthesis L5/S1 level Spondylolysis Spondylosis Spondylosis without myelopathy Sprain of spinal ligament Sprain, lumbosacral ligament Stenosis of intervertebral foramina Stiff back Vertebral osteoporosis Vertebroplasty Wedge fracture of vertebra X-ray of lumbosacral spine	tendon Unstable knee		

Table 3: Included medication names

Simple analgesics (N02BE*)	Non-steroidal anti-inflammatories (M01A*)	Chondroitin and/or glucosamine (M01AX*)	Topical products for joint and muscular pain (M02A*)	Opioids (N02A*)	Other epileptics (N03AX*)
Caffeine, Paracetamol	Celecoxib	Borate, Chondroitin, Glucosamine, Manganese	Benzydamine	**Weak single opioids**	Gabapentin
Paracetamol	Diclofenac	Borate, Chondroitin, Glucosamine, Manganese	Benzydamine hydrochloride	Codeine	Pregabalin
	Diclofenac potassium	Chrondroitin, Copper, Glucosamine, Manganese, Zinc Sulfate	Cajuput oil, Camphor, Capsicum, Eucalyptus oil, Hydroxybenzoate, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil	Codeine phosphate	
Paracetamol combinations	Diclofenac sodium	Chondroitin, Dimethyl Sulfone, Glucosamine	Cajuput oil, Camphor, Clove, Menthol (TIGER BALM)	Codeine phosphate hemihydrate	
Ibuprofen, Paracetamol	Diclofenac sodium, Misoprostol	Glucosamine	Camphor, Menthol, Eucalyptus oil, Methyl salicylate	Dextropropoxyphene	
	Diclofenac, Misoprostol	Glucosamine, Calcium, Vitamin D, Minerals	Camphor, Eucalyptus oil, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil	Dextropropoxyphene napsylate	
	Etoricoxib	Glucosamine, Chondroitin	Camphor, Menthol, Methyl salicylate	Dihydrocodeine	
	Flurbiprofen	Glucosamine hydrochloride	Camphor, Eucalyptus oil, Menthol, Methyl salicylate	Dihydrocodeine tartrate	
	Ibuprofen	Glucosamine hydrochloride, Chondroitin sulphate	Camphor, Eucalyptus oil, Methyl salicylate, Menthol, Alisma plantago aquatica Root oil extract, Bambusa root	Tramadol	
	Ibuprofen lysine	Glucosamine hydrochloride, Chondroitin sulfate, Dimethyl sulfone	Capsaicin	Tramadol hydrochloride	
	Indomethacin	Glucosamine hydrochloride; Chondroitin sulfate, Manganese gluconate, Calcium ascorbate	Capsicum oleoresin, Arnica montana, Arctium lappa root dry, Aloe barbadensis inner leaf juice	**Combination weak opioid**	
	Ketoprofen		Diclofenac	Aspirin, Codeine phosphate	
	Ketorolac			Codeine, Ibuprofen	
	Ketorolac trometamol			Codeine phosphate, Ibuprofen	
	Lumiracoxib			Codeine, Paracetamol	
	Mefenamic acid			Codeine Phosphate, Paracetamol	
	Meloxican			Codeine phosphate hemihydrate, Ibuprofen	
	Naproxen				
	Naproxen sodium				
	Naproxen, Esomeprazole				

	Parecoxib Parecoxib sodium Piroxicam Rofecoxib Sulindac Tiaprofenic acid	Glucosamine hydrochloride, Calcium, Vitamin D, Vitamin K, Boron Glucosamine hydrochloride; Glucosamine sulfate, Glycine, fructose, Bioflavonoids, Ascorbic acid, Histidine, Lysine hydrochloride, Leucine, Valine, Perna caniculata powder, Calcium pantothenate, Zinc amino acid chelate, Manganese amino acid chelate, Copper gluconate, Selenomethionine Glucosamine, Omega-3 triglycerides Glucosamine sulfate Glucosamine sulfate, Chondroitin sulfate (Shark) Glucosamine sulfate, Shark cartilage Glucosamine sulfate, Potassium chloride Glucosamine sulfate sodium chloride, Eicosapentaenoic acid, Docosahexaenoic acid Ascorbate, Glucosamine, Manganese, Turmeric Borate, Glucosamine, Manganese, Selenium	Diclofenac diethylamine Diclofenac diethylammonium Diclofenac Sodium Ethyl salicylate, Hydroxyethyl salicylate, Methyl salicylate, Nicotinic acid Eucalyptus oil Eucalyptus oil, Pine oil Pumilio, Peppermint oil, Camphor, Methyl salicylate, Menthol, Turpentine oil Eucalyptus oil, Menthol, Methyl salicylate Flurbiprofen sodium Ibuprofen Ketoprofen Menthol Menthol, Camphor, Cajuput oil, Clove oil, Dementholised mint oil Menthol, Camphor, Cajuput oil, Dementholised mint oil, Clove bud oil Menthol, Glycol salicylate Menthol, Eucalyptus oil, Methyl salicylate Methyl salicylate Methyl salicylate, Ethyl salicylate, 2-Hydroxyethyl salicylate, Methyl nicotinate	Dextropropoxyphene, Paracetamol Dextropropoxyphene napsylate, Paracetamol Tramadol, Paracetamol Tramadol hydrochloride, Paracetamol **Strong single opioids** Fentanyl Fentanyl citrate Hydromorphone Hydromorphone hydrochloride Morphine Morphine hydrochloride Morphine hydrochloride trihydrate Morphine sulfate Morphine sulfate Bp Morphine sulfate pentahydrate Morphine tartrate Oxycodone Oxycodone, Naloxone Oxycodone hydrochloride Oxycodone pectinate	

		Ascorbate, Cod-liver oil, Colecalciferol, Copper, Cyanocobalamin, Folate, Glucosamine, Manganese, Omega-3 triglycerides, Selenium, Tocopherol, Zinc	Methyl salicylate, Eucalyptus oil, Menthol liquid Methyl salicylate, Menthol Nicoboxil/Nonivamide Nonivamide, Butoxyethyl nicotinate Piroxicam Triethanolamine salicylate Trolamine salicylate	Oxycodone hydrochloride, Naloxone hydrochloride Tapentadol Tapentadol hydrochloride	

* Anatomic and Therapeutic Classifications (ATC) category

Table 4: Test name subgroups for low back, knee, shoulder, and neck imaging tests and procedures

Low back imaging subgroups	Knee imaging subgroups	Shoulder imaging subgroups	Neck imaging subgroups
Lumbosacral plain radiograph*	Knee plain radiograph*	Shoulder plain radiograph*	Neck plain radiograph*
Lumbosacral CT*	Knee CT*	Shoulder ultrasound*	Neck CT*
Lumbosacral MRI*	Knee MRI*	Shoulder MRI*	Neck MRI*
Lumbosacral injection*	Knee injection*	Shoulder injection*	Neck injection*
Lumbosacral unspecified*#	Knee unspecified*#	Shoulder unspecified*#	Neck unspecified*#
Lumbosacral ultrasound^	Knee ultrasound*	Shoulder hydrodilatation*	Neck ultrasound^
Lumbosacral other^	Knee other^	Shoulder other^	Neck other^
	Knee aspiration^	Shoulder aspiration^	Neck aspiration^
	Knee arthrogram^	Shoulder arthrogram^	
		Shoulder CT^	
		Shoulder fluoroscopy^	

* Eligible

^ Ineligible

Analyse as plain radiograph

AUTHOR CONTRIBUTIONS

RH, DOC and RB conceived the study. LB and AG were responsible for data coding and the statistical analysis plan. All authors contributed to refining the protocol and approved the submitted protocol.

FUNDING STATEMENT

This work was supported by an Arthritis Queensland, Arthritis South Australia and the Allan and Beryl Stephens Grant from Arthritis Australia (ID N/A). Arthritis Australia did not contribute to the conduct of this study. It is also supported by an Australian National Health and Medical Research Council (NHMRC) Program Grant (APP1113532). DOC is supported by a TRIP Fellowship and RB is supported by an NHMRC Investigator Fellowship.

COMPETING INTERESTS STATEMENT

RH, DOC, RB and DM report grants from Arthritis Australia (not-for-profit organisation), during the conduct of the study. CP is an employee of Outcome Health, the not-for-profit organisation that developed the POLAR database and chairs the Product improvement group of the Australian Digital Health Agency. It has no relationship with the research but has provided grant funding to Outcome Health. LB reports consultancy fees paid to Monash University from Charite Medical University Berlin, Jesuit Social Services Victoria, and Swinburne University of Technology, outside the submitted work.

REFERENCES

1. Royal Australian College of General Practitioners. General Practice: Health of the Nation. East Melbourne, Vic: RACGP, 2019.
2. Britt H MG, Henderson J, Bayram C, Harrison C, Valenti L, Pan Y, Charles J, Pollack AJ, Wong C, Gordon J. General practice activity in Australia 2015–16. General practice series no. 40. Sydney: Sydney University Press, 2016.
3. Rahman N, Gruber D. The burden of musculoskeletal conditions in Australia: a detailed analysis of the Australian Burden of Disease Study 2011: Australian Institute of Health and Welfare 2017.
4. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020;396(10258):1204-22.
5. Britt H MG, Bayram C, Henderson J, Valenti L, Harrison C, Pan Y, Charles J,, Pollack AJ CT, Gordon J, Wong C. A decade of Australian general practice activity 2006–07 to 2015–16. General practice series no. 41. Sydney: Sydney University Press, 2016.
6. Bennell KL BC, Harrison C, Brand C, Buchbinder R, Haas R, and Hinman RS. Trends in management of hip and knee osteoarthritis in general practice in Australia over an 11-year window: a nationwide cross-sectional survey. *The Lancet Regional Health – Western Pacific* 2021;12 doi: https://authors.elsevier.com/sd/article/S2666606521000961
7. Naunton J, Harrison C, Britt H, et al. General practice management of rotator cuff related shoulder pain: A reliance on ultrasound and injection guided care. *Plos one* 2020;15(1):e0227688.
8. Busingye D, Gianacas C, Pollack A, et al. Data Resource Profile: MedicineInsight, an Australian national primary health care database. *International Journal of Epidemiology* 2019;48(6):1741-41h.
9. Pearce C, McLeod A, Rinehart N, et al. What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program. *Stud Health Technol Inform* 2019;264:303-07.
10. de Lusignan S, Sun B, Pearce C, et al. Coding errors in an analysis of the impact of pay-for-performance on the care for long-term cardiovascular disease: a case study. *Journal of Innovation in Health Informatics* 2014;21(2):92-101.
11. Hemkens LG, Benchimol EI, Langan SM, et al. The reporting of studies using routinely collected health data was often insufficient. *Journal of clinical epidemiology* 2016;79:104-11.
12. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885.
13. Healthmap: Healthdirect Australia; [Available from: https://studio.healthmap.com.au/#/h4f accessed 27th September 2020.
14. The State of Victoria Department of Premier and Cabinet. Population Diversity in Victoria: 2016 Census Local Government Areas, 2018.
15. Yan J, Hawes L, Turner L, et al. Antimicrobial prescribing for children in primary care. *Journal of paediatrics and child health* 2019;55(1):54-58.
16. Turner LR, Pearce C, Borg M, et al. Characteristics of patients presenting to an after-hours clinic: results of a MAGNET analysis. *Australian journal of primary health* 2017;23(3):294-99.
17. Turner LR, Cicuttini F, Pearce C, et al. Cardiovascular disease screening in general practice: General practitioner recording of common risk factors. *Preventive medicine* 2017;99:282-85.
18. Sezgin G, Georgiou A, Hardie R-A, et al. Compliance with pathology testing guidelines in Australian general practice: protocol for a secondary analysis of electronic health record data. *BMJ open* 2018;8(11)
19. Australian Institute of Health and Welfare. Musculoskeletal conditions and comorbidity in Australia. Canberra: AIHW, 2019.

20. Pearce C, McLeod A, Patrick J, et al. Coding and classifying GP data: the POLAR project. *BMJ health & care informatics* 2019;26(1)
21. Cotd R, Rothwell D, Palotay J, et al. The Systematized Nomenclature of Human and Veterinary Medicine-SNOMED International. *Northfield, IL: College of American Pathologists* 1993
22. World Health Organization. WHO collaborating Centre for Drug Statistics Methodology: ATC classification index with DDDs and guidelines for ATC classification and DDD assignment 2020. *Oslo, Norway: Norwegian Institute of Public Health* 2019
23. Hovis KK, Alizai H, Tham S-C, et al. Non-traumatic anterior cruciate ligament abnormalities and their relationship to osteoarthritis using morphological grading and cartilage T2 relaxation times: data from the Osteoarthritis Initiative (OAI). *Skeletal radiology* 2012;41(11):1435-43.
24. Buchbinder R, Underwood M, Hartvigsen J, et al. The Lancet Series call to action to reduce low value care for low back pain: an update. *Pain* 2020;161(1):S57.
25. Mathieson S, Wertheimer G, Maher C, et al. What proportion of patients with chronic noncancer pain are prescribed an opioid medicine? Systematic review and meta-regression of observational studies. *Journal of Internal Medicine* 2020;287(5):458-74.
26. Ittaman SV, VanWormer JJ, Rezkalla SH. The role of aspirin in the prevention of cardiovascular disease. *Clinical medicine & research* 2014;12(3-4):147-54.
27. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology* 2015;44(3):827-36.
28. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PloS one* 2014;9(6):e99825.
29. Australian Bureau of Statistics. Socio-economic indexes for areas (SEIFA) - technical paper 2016. *cat no 20330 55001* 2018
30. Halpin B. SADI: Sequence analysis tools for Stata. *The Stata Journal* 2017;17(3):546-72.
31. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *arXiv preprint arXiv:200300381* 2020
32. Imai C, Hardie R-A, Franco GS, et al. Harnessing the potential of electronic general practice pathology data in Australia: An examination of the quality use of pathology for type 2 diabetes patients. *International Journal of Medical Informatics* 2020;141:104189.
33. Harrison C, Britt H, Miller G, et al. Prevalence of chronic conditions in Australia. *PloS one* 2013;8(7):e67494.

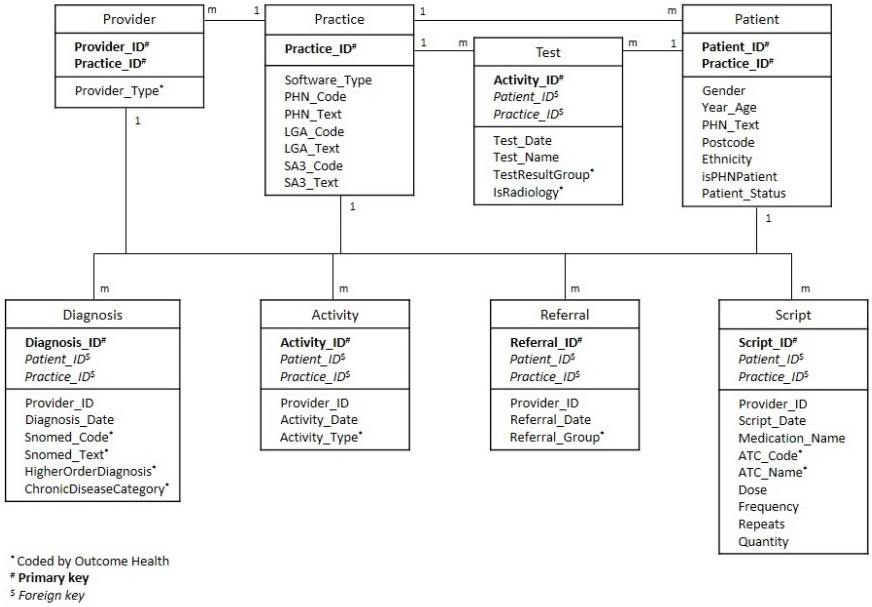

Figure 1

271x186mm (96 x 96 DPI)

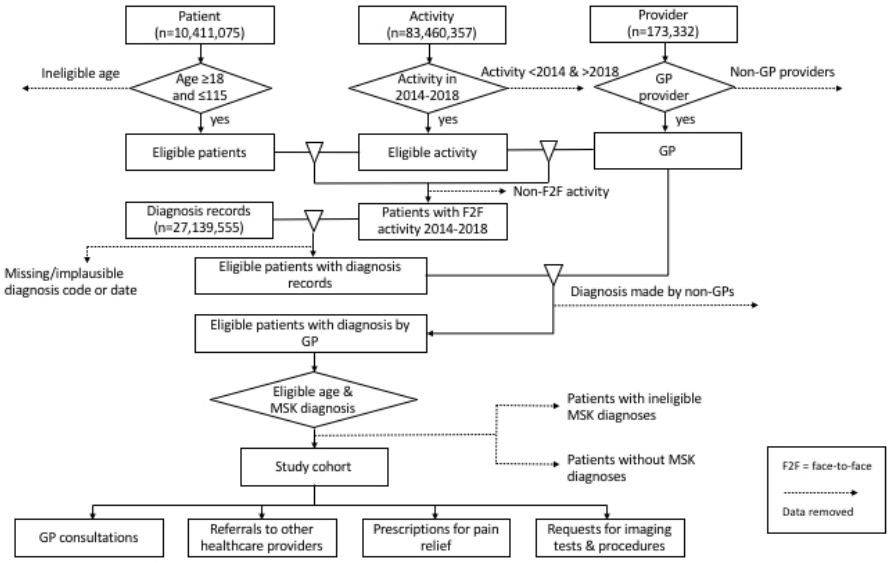

Figure 2

60x34mm (300 x 300 DPI)

Appendix 1. Initial string match terms used to code imaging records

Imaging region strings	X-ray strings	MRI strings	CT strings	Ultrasound strings
Knee				
KNEE	X-RAY KN	MRI KN	CT KN	US KN
KN	X-RAY RIGHT KN	MRI RIGHT KN	CT RIGHT KN	US RIGHT KN
BOTH K*	X-RAY LEFT KN	MRI LEFT KN	CT LEFT KN	US LEFT KN
	XRAY KN	MR KN	KNEE CT	ULTRASOUND KN
Exclude:	XRAY RIGHT KN	MR RIGHT KN		ULTRASOUND RIGHT KN
KNOW	XRAY LEFT KN	MR LEFT KN		ULTRASOUND LEFT KN
KNIGHT	XR KN	MAGNETIC RESONANCE KN		KNEE US
	XR RIGHT KN	MAGNETIC RESONANCE RIGHT KN		KNEE ULT
	XR LEFT KN	MAGNETIC RESONANCE LEFT KN		
	PLAIN FILM KN	KNEE MR		
	PLAIN FILM RIGHT KN	KNEE MAGNETIC		
	PLAIN FILM LEFT KN			
	RADIOGRAPH KN			
	RADIOGRAPH RIGHT KN			
	RADIOGRAPH LEFT KN			
	KNEE X			
	KNEE RADIOGR			
	KNEE PLAIN FILM			
Shoulder				
SHOULDER	X-RAY SH	MRI SH	CT SH	US SH
SH	X-RAY RIGHT SH	MRI RIGHT SH	CT RIGHT SH	US RIGHT SH
CLAVICLE*	X-RAY LEFT SH	MRI LEFT SH	CT LEFT SH	US LEFT SH
	XRAY SH	MR SH	SHOULDER CT	ULTRASOUND SH
Exclude:	XRAY RIGHT SH	MR RIGHT SH		ULTRASOUND RIGHT SH
SHBG	XRAY LEFT SH	MR LEFT SH		ULTRASOUND LEFT SH
TSH	XR SH	MAGNETIC RESONANCE SH		SHOULDER US
SHEET	XR RIGHT SH	MAGNETIC RESONANCE RIGHT SH		SHOULDER ULT
FSH	XR LEFT SH	MAGNETIC RESONANCE LEFT SH		
GSHS	PLAIN FILM SH	SHOULDER MR		

PSH SH/	PLAIN FILM RIGHT SH PLAIN FILM LEFT SH RADIOGRAPH SH RADIOGRAPH RIGHT SH RADIOGRAPH LEFT SH SHOULDER X SHOULDER RADIOGR SHOULDER PLAIN FILM	SHOULDER MAGNETIC		
Neck				
NECK NEC CERVIC C1 C2 C3 C4 C5 C6 C7 C SPINE SPINE CX* **Exclude:** FEMORAL NECK CERVICAL CYTOLOGY	X-RAY NECK XRAY NECK XR NECK X-RAY CERVICAL XRAY CERVICAL XR CERVICAL PLAIN FILM NECK PLAIN FILM CERVICAL RADIOGRAPH NECK RADIOGRAPH CERVICAL NECK X CERVICAL X NECK PLAIN FILM CERVICAL PLAIN FILM NECK RADIOGRAPH CERVICAL RADIOGRAPH	MRI NECK MR NECK MRI CERVICAL MR CERVICAL NECK MR CERVICAL MR	CT NECK CT CERVICAL NECK CT CERVICAL CT	US NECK US CERVICAL ULTRASOUND NECK ULTRASOUND CERVICAL NECK US CERVICAL US NECK ULT CERVICAL ULT
Low back				
LUMB SACR L1 L2 L3 L4 L5	X-RAY LUMB XRAY LUMB XR LUMB X-RAY SACR XRAY SACR XR SACR PLAIN FILM LUMB	MRI LUMB MR LUMB MRI SACR MR SACR LUMBAR MR SACRAL MR	CT LUMB CT SACR LUMBAR CT SACRAL CT	US LUMB ULTRASOUND LUMB US SACR ULTRASOUND SACR LUMBAR US LUMBAR ULT SACRAL US

LOWER BACK **Exclude:** FUNGAL2	PLAIN FILM SACR RADIOGRAPH LUMB RADIOGRAPH SACR LUMBAR X LUMBAR RADIOGRAPH LUMBAR PLAIN FILM SACRAL X SACRAL RADIOGRAPH SACRAL PLAIN FILM			SACRAL ULT

*String match term added after initial coding

Appendix 2

The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	a) Study design (observational cohort study) is included in the title b) Abstract (methods and analysis) contains a summary of what was done. As this is a protocol, what was found is not applicable	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title & abstract Methods and analysis of Abstract refers to geographic region (South Eastern Victoria) and timeframe within the study (1/1/14 to 31/12/18) N/A
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction contains rationale for protocol (explicit reporting of the systematic approach used to classify and select eligible records from the POLAR database will facilitate replication and transparency)		
Objectives	3	State specific objectives, including any prespecified hypotheses	Objectives No prespecified hypotheses reported as this is a protocol		
Methods					
Study Design	4	Present key elements of study design early in the paper	Study design section of Methods (retrospective cohort study)		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Setting section of Methods includes locations of practices with POLAR database, dates of the study period (exposure and data collection) and follow-up		

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	Eligibility criteria are presented in Table 2. Sources and methods of selection are described in Variables section of Methods. Setting – patient-level follow-up data until 31/12/18 will be included N/A	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Variables section of Methods Diagnoses and imaging records within Variables section of Methods N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Variables section of Methods and Tables 3, 4 & 5	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Variables section of Methods and Tables 3, 4 & 5
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Data source section of Methods		
Bias	9	Describe any efforts to address potential sources of bias	Variables and data cleaning sections of Methods		
Study size	10	Explain how the study size was arrived at	Sample size consideration		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Analyses section of Methods		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	a) Analyses section of Methods b) N/A		

		(b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	c) Approach to dataset creation d) Data access and cleaning – only data after which there has been consistent reporting within a practice will be included e) Analyses – sensitivity analysis to include entire follow-up period (instead of 1 year) for each participant and based on date of most recent diagnosis instead of index diagnosis for participants with multiple body regions affected by an eligible musculoskeletal diagnosis		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Data access and cleaning methods Data access and cleaning methods
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	N/A
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	N/A	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	N/A
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on	N/A		

		exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	N/A		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	N/A		
Discussion					
Key results	18	Summarise key results with reference to study objectives	N/A		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	N/A		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion – potential representativeness of the POLAR database and generalizability to the wider population is discussed		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Funding statement		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Appendix 1 – initial string match terms used to code imaging tests

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Patterns of care for people presenting to Australian general practice with musculoskeletal complaints based upon routinely collected data: Protocol for an observational cohort study using the Population Level Analysis and Reporting (POLAR) database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055528.R1
Article Type:	Protocol
Date Submitted by the Author:	06-Sep-2021
Complete List of Authors:	Haas, Romi; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine Busija, Ljoudmila; Monash University, Department of Epidemiology and Preventive Medicine Gorelik, Alexandra; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine O'Connor, Denise; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine Pearce, Christopher; Outcome Health Mazza, Danielle; Monash University, Department of General Practice Buchbinder, Rachelle; Cabrini Health, Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology; Monash University, Department of Epidemiology and Preventive Medicine
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Rheumatology
Keywords:	PRIMARY CARE, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, Back pain < ORTHOPAEDIC & TRAUMA SURGERY, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Knee < ORTHOPAEDIC & TRAUMA SURGERY, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Patterns of care for people presenting to Australian general practice with musculoskeletal complaints based upon routinely collected data: Protocol for an observational cohort study using the Population Level Analysis and Reporting (POLAR) database

Romi Haas (romi.haas@monash.edu) [1,2*], Ljoudmila Busija (lucy.busija@monash.edu) [2], Alexandra Gorelik (Alexandra.gorelik@monash.edu) [1,2], Denise O'Connor (denise.oconnor@monash.edu) [1,2], Christopher Pearce (drchrispearce@mac.com) [3], Danielle Mazza (danielle.mazza@monash.edu) [4], Rachelle Buchbinder (rachelle.buchbinder@monash.edu) [1,2]

[1] Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Health, Malvern, Victoria, Australia.
[2] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia.
[3] Outcome Health, Blackburn, Victoria, Australia.
[4] Department of General Practice, Monash University, Notting Hill, Victoria, Australia.
[*] Corresponding author

Correspondence

Romi Haas
Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Institute
154 Wattletree Rd, Malvern
Victoria 3144
Australia
Email: romi.haas@monash.edu
Telephone: +61 3 9508 3499

Keywords
Protocol
Primary care
Musculoskeletal disorders
Low back pain
Neck pain
Shoulder pain
Knee pain

ABSTRACT

Introduction

General practice is integral to the Australian healthcare system. Outcome Health's POpulation Level Analysis and Reporting (POLAR) database uses de-identified electronic health records to analyse general practice data in Australia. Previous studies using routinely collected health data for research have not consistently reported the codes and algorithms used to describe the population, exposures, interventions and outcomes in sufficient detail to allow replication. This paper reports a study protocol investigating patterns of care for people

presenting with musculoskeletal conditions to general practice in Victoria, Australia. Its focus is on the systematic approach used to classify and select eligible records from the POLAR database to facilitate replication. This will be useful for other researchers using routinely collected health data for research.

Methods and analysis

This is a retrospective cohort study. Patient-related data will be obtained through electronic health records from a subset of general practices across three primary health networks (PHN) in South Eastern Victoria. Data for patients with a low back, neck, shoulder and/or knee condition and who received at least one GP face-to-face consultation between 1/01/2014 and 31/12/2018 will be included. Data quality checks will be conducted to exclude patients with poor data recording and/or non-continuous follow-up. Relational data files with eligible and valid records will be merged to select the study cohort and the GP care received (consultations, imaging requests, prescriptions and referrals) between diagnosis and 31/12/2018. Number and characteristics of patients and GPs, and number, type and timing of imaging requests, prescriptions for pain relief and referrals to other health providers will be investigated.

Ethics and dissemination

Ethics approval was obtained from the Cabrini and Monash University Human Research Ethics Committees (Reference Numbers 02-21-01-19 and 16975 respectively). Study findings will be reported to Outcome Health, participating PHNs, disseminated in academic journals and presented in conferences.

ARTICLE SUMMARY

Strengths and limitations of this study

- This is the first study to our knowledge to report the codes and algorithms used to classify, select and merge eligible records from the POLAR database into a patient-centred database to facilitate analysis of general practice patterns of care.
- The systematic approach used in this study can be adapted by other researchers using routinely collected health data for research purposes.
- This study will extend previous research that has assessed the representativeness of POLAR data to GP care across the wider Australian population.
- These data are likely to underestimate actual allied health visits as some of these do not require a GP referral in Australia; some prescriptions for pain relief are available without a prescription so these data will also be underestimated.
- It is possible not all patterns of care for the study cohort will be directly attributable to a musculoskeletal condition as reasons for GP consultations, referrals, and prescriptions are not mandated by the source Electronic Medical Records (EMRs).

INTRODUCTION

General practice plays an essential role in providing primary health care to the population. In Australia 86% of the population visits a general practitioner (GP) multiple times a year[1], and nearly 20% of these consultations are for a musculoskeletal condition[2]. These conditions account for 23% of the years lived with disability in Australia[3] and are also a major cause of disability worldwide[4]. Until 2016, the BEACH (Bettering the Evaluation and Care of Health) program provided the most comprehensive data on clinical activities of Australian general practice[5]. The program identified a number of activities that represent low-value care for people with musculoskeletal conditions including an over-reliance on imaging, prescription of opioids, and unnecessary referrals to specialist care[6 7]. However, in-depth exploration of these activities within the BEACH program is limited by its cross-sectional design, and these data are no longer being collected.

Technological advancements have facilitated the extraction of de-identified patient information from general practice clinical information systems. The advantage of these datasets for research purposes are that they are longitudinal and can therefore be used to establish sequences of events at the patient level and to examine changes in GP management over time. Both the Medicine Insight[8] and the Population Level Analysis and Reporting (POLAR) databases[9] are examples of longitudinal general practice datasets within Australia. Unlike POLAR, the Medicine Insight program does not currently include referrals provided by GPs to other healthcare providers[8]. These data may provide important insights into how well GPs are playing their role as gatekeepers of the Australian healthcare system.

While using routinely collected data for research purposes offers considerable opportunities to improve health care, there are several challenges to be overcome. Differences in patient information management and data extraction tools result in variability in both the information captured and ways in which this information is coded. In particular, the way in which text values (diagnoses, examination findings, test results and medications) are transformed to codes can be a source of variation within and between studies. Previous studies have highlighted how code selection affects the reported prevalence and precision of results[10]. Studies conducted using routinely collected health data should therefore be reported with sufficient detail and clarity to allow replication. However, a systematic evaluation of a random sample of 124 publications using routinely collected health data has demonstrated inadequate reporting of the methods used[11]. For example, in 44 studies where definitions of codes or classification algorithms were deemed necessary to describe the population, exposures or interventions and outcomes, only nine (20.5%) reported all three items adequately. The REporting of studies Conducted using Observational Routinely collected Data (RECORD) guidelines, published in 2015, were developed to assist in this process and to ensure that readers can assess the internal and external validity of the findings of these studies[12].

The POLAR database draws data from every consultation occurring for millions of patients in approximately 30% of general practices across South-Eastern Victoria[13], an area that comprises more than half of Victoria's population[14]. Inclusion is based on practice consent so this volume is increasing exponentially as more practices consent to add their data and as more consultations occur over time. Unlike in other countries, coding is not embedded in the clinical process and needs to be conducted specifically for research purposes. Data are provided to research users in a relational database that organises data into files that can be merged based on common data fields. Identifying and selecting relevant records and merging separate files

into a patient-centred database for analysis is a complex task that could potentially yield variable results depending on the methods used.

Previous studies have used the POLAR database to investigate patterns of antimicrobial prescribing for children[15], to examine characteristics of patients presenting to an after-hours clinic[16], to estimate GP recording of cardiovascular risk factors[17], and to describe characteristics of pathology test ordering in general practice[18]. However, these studies have not reported the methods used to classify and select eligible records or the processes used to merge data files into a patient-centred database for analysis.

This manuscript presents a protocol for a study investigating patterns of GP care for people with a low back, neck, shoulder and/or knee condition in Victoria, Australia. It describes the methods used to classify and select eligible records from the POLAR database and how relational data files will be merged into a patient-centred database. This systematic approach will guide future research by enabling researchers interested in using routinely collected health data, and the POLAR database in particular, to answer other clinically relevant questions about general practice care. Study findings will advance existing knowledge about GP care for people with these musculoskeletal conditions and whether it conforms to best evidence-based practice. Differences in care across different musculoskeletal complaints may also inform tailored interventions to improve care and ultimately reduce the burden of disease associated with these musculoskeletal complaints.

Objectives

The aim of this study will be to examine GP patterns of care for people with low back, neck, shoulder and knee conditions. Specific objectives will be to:

1. Describe and compare the management (number, type and timing of imaging tests and procedure requests, prescriptions for pain relief, and referrals to other health providers) provided by GPs to people with low back, neck, shoulder and knee conditions
2. Describe the prevalence of comorbidities among specific musculoskeletal diagnoses within this cohort
3. Examine the association between management types and patient- and practice-related variables
4. Examine the longitudinal changes in GP management for these conditions between 2014 and 2018 inclusive

METHODS

Study design

A retrospective cohort study using general practice health records from Victoria, Australia.

Data source

This study will use data from Outcome Health's POLAR database[9]. The database structure is based on eight relational files, each containing de-identified practice, provider, and/or patient codes (Figure 1). These common fields allow merging of the data files so that databases can be configured for specific research purposes. Data is extracted from two different clinical information systems, covering ninety percent of included general practices. All data is extracted using the Hummingbird data extraction tool[9].

Setting

The POLAR database contains de-identified patient-related data from all electronic medical records of consenting general practices within the PHNs of Eastern Melbourne, South Eastern Melbourne and Gippsland within Victoria, Australia. Our study will include data collected over five calendar years from 1 January 2014 until 31 December 2018 relating to all patients with an eligible musculoskeletal condition and who received at least one face-to-face GP consultation. Follow up will be from the time of the initial recorded diagnosis to 31st December 2018. Data analyses will be completed by the end of 2021.

Participants

The study cohort will include people diagnosed during 2014 to 2018 inclusive with a low back, neck, shoulder and/or knee condition, limited to age 45 years and over except for low back which will be limited to age 18 years and over. The differing age restrictions were chosen because the prevalence of most musculoskeletal conditions increases markedly after the age of 45 except for low back pain which increases after the age of 18[19]. Eligibility criteria are presented (Table 1). We excluded traumatic diagnoses and conditions typically primarily managed by a specialist (e.g., inflammatory and autoimmune rheumatic diseases). Patients with an eligible diagnosis and age will also have received at least one GP face-to-face consultation during the study dates. The musculoskeletal diagnosis will not have to occur during a GP consultation since it is an eligible diagnosis that could result from consultation with other health care providers.

Variables

Preparatory work to classify and select eligible records has been completed as part of the protocol process. In circumstances where Outcome Health has previously coded data (e.g., diagnosis records), we used this coding to select eligible records that fitted our inclusion criteria. In circumstances where there was no coding (e.g., imaging tests), we coded the data into categories and then selected eligible records. Outcome Health's approach to coding used clinical natural language processing to automatically code structured narrative text within the electronic medical record following by a manual process for quality checking and correction[20]. For example, this allowed the free text items 'back pain', 'low back pain', and 'lumbar pain' to all sit under the same diagnostic code. Where possible, coding was conducted using a standardised classification system. For example, diagnoses are coded using SNOMED CT-AU terminology[21] and prescriptions are coded according to the Anatomical Therapeutic Chemical (ATC) classification system[22]. In cases where there is no standardised classification system available (e.g., providers and referrals), Outcome Health used a similar process to code these variables into relevant categories (e.g., type of health care provider). Clinical natural language processing conducted by Outcome Health has previously demonstrated accurate coding of over 95% of the narrative text to SNOMED CT-AU terms in a sample of approximately 57,000 diagnosis records[20]. Our approaches to coding and/or selecting eligible records for each variable are described in detail below.

Provider records

Healthcare providers other than a GP may be nested within a general practice. To limit all diagnoses, consultations, referrals, and prescriptions to those made only by GPs we used coding within the provider type field conducted by Outcome Health. This is coded by Outcome Health

according to the professional background of the healthcare provider delivering the service (e.g., GP, nurse).

Diagnoses records

All SNOMED CT-AU diagnosis-related terms used during 2014-2018 were searched by two study authors (RH and RB) to select eligible low back, neck, shoulder and knee conditions. We included all patients with an eligible musculoskeletal diagnosis during 2014-2018 regardless of whether they had a prior musculoskeletal diagnosis. Included SNOMED diagnosis terms are presented (Table 2). Sacral conditions were included as part of low back conditions. The following SNOMED terms were excluded as these conditions were deemed to be indicative of traumatic injury or conditions that are not managed primarily by GPs: fracture (except lumbar and tibial plateau fractures), dislocation, synovectomies/synovitis, and cauda equina syndrome. Knee ligamentous and meniscal tears were included as these are likely due to degeneration in the 45 years and over age group[23]. Lesions were excluded as these could involve a wound, ulcer or tumour and are not musculoskeletal conditions. General musculoskeletal terms such as sprain or osteoarthritis (where the site was not specified) were also excluded as these could not be attributed to a specific body region. We included relevant surgical or procedural musculoskeletal terms as GPs are involved in referral and follow-up for these conditions.

Using experienced clinicians, Outcome Health has further categorised SNOMED diagnoses into overarching groups and utilised key chronic disease groups as a qualifier[9]. For example, free text such as 'low back pain' or 'angina' could be qualified as a chronic disease if present for six months or more. We used these chronic disease groups to identify eligible comorbid diagnoses for our study cohort as follows: chronic cardiovascular disease, chronic obstructive pulmonary disease, chronic musculoskeletal conditions, cancer, opioid addiction, dementia, diabetes, depression/anxiety, and obesity. Obesity was identified using SNOMED terms as it was not coded as a chronic disease category in the POLAR database. We included previous chronic musculoskeletal conditions so that these could be investigated as a potential predictor of different management patterns.

Activity records

Activity records are coded in POLAR according to the type of consultation provided (e.g., telehealth, visit, telephone). Each time a note is recorded in the narrative section it is coded by the EMR and this is extracted by POLAR. We used this coding to select eligible patients who had at least one 'Activity type' relating to a face-to-face consultation (i.e., encounter, surgery or visit) during 2014-2018 inclusive. Telehealth and telephone consultations were also included for follow-up consultations only.

Referral records

Referral records are coded in POLAR according to discipline (e.g., neurosurgeon, physiotherapist, endocrinologist). We used this coding to select eligible referral groups considered relevant to a person with low back, neck, shoulder or knee conditions. The following referral groups were included: orthopaedics and neurosurgery (surgical specialists); sports medicine, rheumatology, rehabilitation medicine, neurology, and pain management (non-surgical specialists); and physiotherapy, osteopathy, massage therapy, exercise physiology, chiropractor, and psychology (allied health providers).

Prescription records

Medications are coded in POLAR according to the Anatomic and Therapeutic Classifications (ATC) system[22]. We included medications deemed by the study authors to be commonly prescribed for pain relief to people with musculoskeletal conditions. Medications within the following categories were included: simple analgesics; non-steroidal anti-inflammatories (NSAIDs); chondroitin and/or glucosamine; topical products for joint and muscular pain; opioids; neuromodulators and any relevant combinations. We included neuromodulators such as gabapentin and pregabalin because these are being increasingly used for the management of musculoskeletal conditions such as nonspecific low back pain or sciatica despite evidence of a lack of effectiveness and a higher risk of adverse events[24]. Opioid analgesics were further categorised into (i) weak single ingredient opioid analgesics (e.g. codeine), defined as <50 morphine milligram equivalents (MME) per day; (ii) strong single ingredient opioid analgesics (e.g. tapentadol, oxycodone, morphine), defined as 150 MME per day; and (iii) combination opioid analgesics[25]. Medicines in the combination opioid category were categorised based on the strongest medicine present, either as a weak combination opioid or as a strong combination opioid.

To ensure we included all potentially eligible medication names, we searched by both ATC category and by medication name from the prescription file during 2014-2018. The medication names we included are presented in Table 3. We included oral, topical and injectable preparations of medications. We excluded the following prescriptions: aspirin, decongestants (e.g., pseudoephedrine), antihistamines (e.g., doxylamine), opioid cough suppressants (e.g., dextromethorphan), and expectorants (e.g., guaifenesin). These were excluded on the basis that they were likely to have been prescribed for another condition (e.g., aspirin for secondary prevention of cardiovascular disease[26]).

Imaging records

The test data file within POLAR contains radiology and pathology tests requested by the GP. At the time of data extract, coding of the test data file had not been completed for specific imaging tests by Outcome Health and there were too many records to scan manually. We therefore exported all radiology test names during 2014-2018 inclusive and used an inductive coding process to select the following eligible imaging tests: plain radiographs, CT and MRI scans of the lumbar and cervical spine; plain radiographs, CT, MRI and ultrasounds of the knee; and plain radiographs, MRI scans and ultrasounds of the shoulder. We also included lumbar spine, knee, shoulder and cervical spine injections and shoulder hydrodilatation as eligible radiology procedures.

To code eligible imaging records, we first used the string match command in Stata to select all test names for each eligible anatomical region (i.e., low back, neck, shoulder and knee). Within each region, we then iteratively coded all imaging records into subgroups according to the type of imaging test (e.g., ultrasound). This process involved developing string match terms to identify each type of eligible radiology test or procedure within the sample, reviewing the uncoded test names (subgrouped as 'other') and manually coding additional terms until the remaining test names could not be classified into any further subgroups. We also developed string match terms to identify bilateral tests of the shoulder and knee. The initial string match terms used to code each body region and eligible imaging test or procedure are presented in Appendix 1.

During the coding process, there were numerous test names that did not definitively identify a type of imaging test (e.g., 'right knee'). We labelled these as 'unspecified'. We plan to classify these as plain radiographs in our analysis. This is because plain radiograph was deemed to be the default radiology modality in the Electronic Medical Record (EMR) software. The subgroups of imaging records inductively developed for each eligible body region are presented in Table 4. Our subgroup coding (excluding test names labelled as 'unspecified' and 'other') accounted for 96%, 95.8%, 95.2% and 96.6% of the identified low back (n=180,630), neck (n=192,844), shoulder (n=236,803) and knee (n=235,123) imaging test names respectively.

Test names indicating more than one imaging test were classified separately. We excluded imaging tests of soft tissues of the neck and test names indicating a combined neck image with the head, larynx, thyroid and/or abdomen (unless it specifically stated cervical spine) as we deemed these investigations were most likely not requested for a musculoskeletal condition. We also excluded test names with the following terms as these were not deemed to indicate an imaging test or procedure: 'report', 'findings', 'cancel', 'results', 'letter'.

Data access and cleaning

Outcome Health provided the research team with access to all POLAR database records since inception (1997). Data quality checks will be performed to label data as 'acceptable' for analysis using a similar process to that conducted by an established general practice database in the United Kingdom[27]. Duplicate data and records with empty or implausible birth dates (defined as greater than 115 years of age at time of diagnosis or dated after patient management) will be excluded from analyses. We will exclude practices without any activity data during 2014-2018. We will also examine the consistency of activity, test, prescription, and referral data for each practice in each eligible calendar year. If a gap in reporting from any practice is identified for one year or more, only data from the earliest date after which there was no gap will be included. For example, if a practice has activity data in 2014, 2017 and 2018, only data from 2017 onwards will be included. In addition, we will exclude activity records that represent more than one face-to-face consultation with a GP for the same patient on the same day. This is because an 'activity' occurs in POLAR anytime a patient record is accessed regardless of whether this was for clinical or administration purposes.

Approach to dataset creation

We will use a systematic process to systematically exclude ineligible records in order to merge data and select the study cohort (Figure 2). This process will require the merging of five relational data files (patient, practice, provider, activity and diagnosis) in a specific sequence to ensure all relevant records are retained. For example, we will not limit diagnosis records to 2014-2018 until after we have selected relevant comorbidities. A patient-centred database will be prepared to examine the number and type of GP consultations, imaging test and procedure requests, prescriptions for pain relief, and referrals to other health providers for our study cohort. Data that does not match our eligibility criteria (including data with missing fields) will be excluded during the merging process as unmatched records. Duplicate records, records with implausible dates or missing fields, and multiple records of the same type on a single day will also be removed and reported.

Analyses

All relevant data will be extracted from the POLAR SQL database and imported into Stata 15 (STATA Corp LP, College Station, TX, USA) for data management and analyses. The methods in this protocol are structured according to RECORD guidelines (Appendix 2)[12]. Full lists of codes used to define eligible variables are available from https://clinicalcodes.rss.mhs.man.ac.uk/medcodes/article/174/[28] where there is a recognised coding system.

Descriptive statistics will be used to summarise the study cohort including the number and type of eligible musculoskeletal conditions, patient demographics and comorbidities. These will be compared to national health survey data to assess the representativeness of the POLAR database to the wider Australian population. Eligible musculoskeletal conditions will be grouped according to body region.

Primary analysis will include analysis of each management type provided for each participant during the first year after their index diagnosis. A sensitivity analysis will be conducted including the entire follow-up period until 31st December 2018. For prescriptions, the primary analysis will include the entire follow-up period because repeated prescriptions over more than one year are anticipated. Descriptive statistics will also be used to summarise the number and type of GP all-cause consultations, imaging tests and procedures requested, prescriptions for pain relief, and referrals to other health providers for the study cohort. Results will be stratified by affected body region. Consultations will be categorised as face-to-face or telecommunication. Imaging requests will be categorised according to the type of imaging modality or procedure and body region (e.g., knee MRI). Bilateral knee and shoulder imaging requests will be counted as two imaging requests. Prescriptions will be categorised according to paracetamol, NSAIDs, glucosamine and/or chondroitin, opioids (weak single opioid, strong single opioid, weak combination opioid, and strong combination opioid) and neuromodulators. Referrals will be categorised according to surgical specialist, non-surgical specialist, and allied health. Patterns and timing of management (imaging requests, prescriptions and referrals) for people with eligible low back, neck, shoulder and knee conditions will be examined and compared between each year within the five-year study period and relative to time of diagnosis using trend analyses.

One of the limitations of the POLAR database is that it does not capture reasons for the clinical encounter or management types (imaging request, prescription or referral). To account for the subsequent uncertainty in attributing management types to a particular diagnosis for those with multiple musculoskeletal conditions, participants with eligible musculoskeletal diagnoses from multiple body regions will be analysed separately to those with eligible diagnoses in one body region. For participants with multiple eligible musculoskeletal diagnoses throughout the study period, the primary analysis will be conducted relative to the date of the first (index) eligible musculoskeletal diagnosis and a sensitivity analysis relative to the date of the most recent diagnosis will also be conducted. Imaging requests will be analysed relative to the date of the most recent musculoskeletal diagnosis for the same body region. For example, a shoulder ultrasound will be analysed

relative to the index date of an eligible shoulder diagnosis even if the same patient was diagnosed previously with an eligible knee condition.

The association between management types and patient- and practice-related characteristics will be examined using regression analysis. Predictors will include patient age, gender, body region(s) affected by eligible musculoskeletal conditions, socioeconomic status, remote or metropolitan location of GP practice, whether the patient lives within the Primary Health Network (PHN) of the practice, and time since diagnosis. Socioeconomic status will be defined by the Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD) using 2016 Census data[29].

Sequence analysis will be used to categorise sequences of management types of people with eligible musculoskeletal conditions into similar groups based on observed characteristics[30]. This will take into account both the time since diagnosis and sequence of each management type. We will use this to identify the most frequently used combinations and sequences of management and the patient- and practice-related variables that correlate with each management combination.

Sample size consideration

Sequence analysis will require the largest sample size of our planned analyses and will therefore form the basis of our sample size consideration. We plan to examine the following six management types: non-surgical referrals, surgical referrals, allied health referrals, opioid prescription, X-ray and/or ultrasound requests, and MRI and/or CT scan requests. This provides a total of 720 potential sequence combinations. Based on a recommended 20 to 30 subjects per subgroup[31], we estimate a sample size of between 14,400 and 21,600 will be required to differentiate between each sequence combination or pattern of care. Recent use of the POLAR database using data from approximately 200 general practices identified 20,514 active adult patients with type 2 diabetes before July 2016[32]. Our extract is based on 301 general practices from 2014 to 2018 and since the prevalence of diabetes is less than that of musculoskeletal conditions[33], we expect a sample size of more than 20,000.

Patient and public involvement

There will be no involvement of patients or the public in this study.

DISCUSSION

Explicitly reporting our systematic approach used to classify, select and merge eligible records from relational data files into a patient-centred database for analysis promotes transparency, reproducibility and completeness of the reporting of research conducted using routinely collected health data. The approach used to code eligible imaging tests from structured narrative text coded over 95% of the 845,400 cumulative imaging-related test and procedure records identified for low back, shoulder, knee and neck conditions during 2014-2018. Our code lists are available for all variables that have been previously coded by POLAR and those with a recognised coding system have been made available on the ClinicalCodes online repository. Although our coding process may only be applicable to systems that do not embed

coding in the clinical process, this approach can also be adapted to examine patterns of care over time for other conditions in general practice.

The main strength of this study is that it will facilitate an overview of the care provided by GPs to the same patient(s) over time and thereby enable temporal sequences to be examined. The POLAR database contains all patient-related activity within each practice making it representative of the included practices. Previous research has demonstrated comparable prevalence and age-gender distribution of people diagnosed with type 2 diabetes within the POLAR database to those within Australia[32]. This study will add to these findings by assessing the representativeness of people with musculoskeletal conditions within the POLAR database to the wider Australian population.

Constraints within the POLAR database may potentially limit the reliability of this study's findings although these are problems inherent in the use of any extracted data. Variability in workflows and recording behaviour introduces potential biases and the different clinical information systems used by the practices within POLAR may result in variability in the information entered. The objective of POLAR is to remove as much variability as possible by using and being transparent about the coding process. High accuracy of diagnostic coding by Outcome Health has been previously demonstrated[20]. In addition, it is possible not all patterns of care for the study cohort will be directly attributable to a musculoskeletal condition because reasons for GP consultations, referrals, and prescriptions are not mandated in the source EMRs. Prescriptions for some types of pain relief and referrals to allied health providers may also be underestimated by the POLAR database as these forms of management may be generated outside of general practice. These data are also likely to underestimate actual allied health visits and prescriptions for pain relief as some of these do not require a GP referral and are available over-the-counter without a prescription respectively in Australia.

ETHICS AND DISSEMINATION

Prior approval to conduct this study was obtained from the Cabrini Human Research Ethics Committee and Monash University Human Research Ethics Committee (Reference Numbers 02-21-01-19 and 16975 respectively). We did not obtain participant consent as all data was anonymised. Outcome Health holds a standing ethics approval for its collection and custodianship of the data from the Royal Australian College of General Practice. The study findings will be reported to Outcome Health, participating PHNs, disseminated in peer-reviewed academic journals and presented in national and international conferences.

FIGURES

Figure 1. Database structure
Figure 2. Approach to dataset creation

Table 1. Eligibility Criteria

Patient population			Patient management				
Diagnoses	**Provider**	**Patient**	**Practice**	**Activity**	**Referrals**	**Prescriptions**	**Imaging tests & procedures**
Low back Knee Shoulder Neck **Exclude:** Trauma Systemic inflammatory arthritis	Diagnosed by a general practitioner	Aged ≥18 years for low back conditions Aged ≥45 years for all other diagnoses	Patient activity 2014-2018	Face-to-face Telehealth	Surgical specialists Non-surgical specialists Allied health providers e.g. psychologist	Simple analgesics Anti-inflammatories Chondroitin/ Glucosamine Topical products Opioids Neuromodulators	Lumbar plain radiograph Lumbar CT Lumbar MRI Lumbar injection Knee plain radiograph Knee CT Knee MRI Knee ultrasound Knee injection Shoulder plain radiograph Shoulder ultrasound Shoulder MRI Shoulder injection Shoulder hydrodilatation Cervical plain radiograph Cervical CT Cervical MRI Cervical injection

Table 2: Included SNOMED terms

Low back diagnoses	Knee diagnoses	Shoulder diagnoses	Neck diagnoses
Arthritis of spine	Acute meniscal tear, medial	Acromioclavicular joint structure	Cervical arthritis
Arthropathy of spinal facet joint	Anterior knee pain	Adhesive capsulitis of shoulder	Cervical arthrodesis
Back problem	Arthritis of knee	Arthritis of acromioclavicular joint	Cervical disc disorder
Backache	Arthrodesis of knee	Arthrodesis of shoulder	Cervical kyphosis
Bone structure of coccyx	Arthroscopic lateral patellar release	Arthrography of shoulder	Cervical laminectomy
Bone structure of L5	Arthroscopic meniscectomy	Arthroscopic acromioplasty	Cervical myelopathy
Bone structure of sacrum	Arthroscopic procedure	Arthroscopic shoulder decompression	Cervical nerve root compression
Chondrectomy of spine	Arthroscopy of knee	Arthroscopy of shoulder	Cervical radiculitis
Chronic back pain	Arthroscopy of knee with lateral meniscectomy	Bursitis of shoulder	Cervical radiculopathy
Chronic lower back pain	Arthroscopy of knee with medial meniscectomy	Calcific tendinitis	Cervical rib
Compression fracture	Arthrotomy of knee	Calcific tendinitis of shoulder	Cervical spinal fusion by anterior technique
Compression fracture of vertebral column	Aspiration of knee joint	Capsulitis	Cervical spine degeneration
Compression of lumbar nerve root	Both knees	Contusion of shoulder region	Cervical spine structure
Correction of scoliosis	Bursitis of knee	Detachment of the glenoid labrum and/or capsule of the shoulder joint	Cervicogenic headache
Crush fracture of lumbar vertebra	Calcium pyrophosphate deposition disease	Entire tendon of supraspinatus muscle	Cervico-occipital neuralgia
CT of lumbar region	Chondrocalcinosis	Full thickness rotator cuff tear	Chronic neck pain
CT of lumbar spine	Chondromalacia of patella	Impingement syndrome of shoulder region	CT of cervical spine
CT of spine	Complete tear, knee, medial collateral ligament	Inflammation of rotator cuff tendon	CT of neck
Curvature of spine	Contusion of knee	Injury of glenoid labrum of shoulder joint	Degeneration of cervical intervertebral disc
Decompression laminectomy	Derangement of knee	Injury of shoulder region	Diffuse cervicobrachial syndrome
Decompression of lumbar spine	Disorder of patellofemoral joint	MRI of shoulder	Excision of cervical intervertebral disc
Degeneration of intervertebral disc	Finding of tear meniscus	Osteoarthritis of acromioclavicular joint	Injury of cervical spine
Degeneration of lumbar intervertebral disc	Fracture of tibial plateau	Osteoarthritis of shoulder	Kyphoscoliosis deformity of spine
Diagnostic radiography of coccyx	Haemarthrosis of knee	Painful arc syndrome	Kyphosis deformity of spine
Discitis	Inflammation of bursa of patella	Radiography of shoulder	Magnetic resonance imaging of neck
Discogenic pain			MRI of cervical spine

Disorder of joint of spine	Injury of anterior cruciate ligament	Repair of musculotendinous cuff of shoulder	Muscle spasm of cervical muscle of neck
Disorder of vertebra	Injury of knee	Repair of shoulder	Neck injury
Exploration of spine	Knee joint - varus deformity	Rotator cuff impingement syndrome	Neck pain
Facet joint pain	Knee joint effusion	Rotator cuff syndrome	Neck sprain
Fracture of body of vertebra	Knee joint valgus deformity	Rupture of tendon of biceps	Neck structure
Fracture of lumbar spine	Knee locking	Rupture of tendon of biceps, long head	Pain in cervical spine
Fracture of sacrum	Knee pain	Shoulder pain	Prolapsed cervical intervertebral disc
Fracture of vertebral column	Knee region structure	Shoulder reconstruction	Radiography of cervical spine
Injury of back	Knee stiff	Shoulder region structure	Spinal stenosis in cervical region
Injury of coccyx	Loose body in knee	Shoulder strain	Stiff neck
Intervertebral disc prolapse	MRI of knee	Shoulder tendinitis	Strain of neck muscle
L4/5 disc	Osteoarthritis of knee	Sprain of acromioclavicular ligament	Strain of tendon of neck
L5/S1 disc	Osteotomy of proximal tibia	Sprain of shoulder	Torticollis
Laminectomy	Osteotomy of tibia	Structure of left shoulder region	Whiplash injury to neck
Lordosis deformity of spine	Patellar instability	Structure of right shoulder region	
Low back pain	Patellar maltracking	Structure of rotator cuff including muscles and tendons	
Low back strain	Patellar tendonitis	Subacromial bursitis	
Lower back injury	Patellectomy	Subdeltoid bursitis	
Lower back structure	Patellofemoral osteoarthritis	Subluxation of acromioclavicular joint	
Lumbar	Patellofemoral stress syndrome	Subscapularis tendinitis	
Lumbar discectomy	Prepatellar bursitis	Supraspinatus tear	
Lumbar laminectomy	Problem knee	Supraspinatus tendinitis	
Lumbar microdiscectomy	Radiologic examination of knee	Total shoulder replacement	
Lumbar radiculopathy	Repair of anterior cruciate ligament of knee joint	US shoulder region	
Lumbar region back structure	Repair of knee collateral ligaments		
Lumbar spinal fusion	Repair of knee cruciate ligaments		
Lumbar sprain	Repair of meniscus		
Lumbosacral spine	Repair of patellar tendon		
Lumbosacral spondylosis	Replacement of total knee joint		
Lumbosacral spondylosis without myelopathy	Rupture of anterior cruciate		

Lumbosacral strain Lumbosacral radiculopathy Magnetic resonance imaging of spine Manipulation of spine MRI of lumbar spine Nerve root compression syndrome Nerve root disorder Operative procedure on spinal structure Osteoarthritis of lumbar spine Pain in lumbar spine Pain in the coccyx Prolapsed lumbar intervertebral disc Radiography of spine Sacral back pain Sacroiliac arthrodesis Sacroiliac joint inflamed Sacroiliac joint pain Scoliosis deformity of spine Scoliosis of lumbar spine Spasm of back muscles Spinal arthritis deformans Spinal arthrodesis Spinal claudication Spinal injury Spinal stenosis Spinal stenosis of lumbar region Spondylitis Spondylolisthesis	ligament Rupture of cruciate ligaments Rupture of medial collateral ligament of knee Rupture of posterior cruciate ligament Sprain of knee Sprain of lateral collateral ligament of knee Sprain of medial collateral ligament of knee Stabilisation of patellofemoral joint Strain of knee Strain of patellar tendon Strain of tendon of medial thigh muscle Structure of left knee Structure of prepatellar bursa Structure of right knee Subluxation of patellofemoral joint Suprapatellar bursitis Swollen knee Synovial cyst of knee Synovial cyst of popliteal space Tear of lateral meniscus of knee Tear of medial meniscus of knee Tear of meniscus of knee Total knee replacement Total replacement of left knee joint Total replacement of right knee joint Traumatic rupture of patellar		

Spondylolisthesis L5/S1 level Spondylolysis Spondylosis Spondylosis without myelopathy Sprain of spinal ligament Sprain, lumbosacral ligament Stenosis of intervertebral foramina Stiff back Vertebral osteoporosis Vertebroplasty Wedge fracture of vertebra X-ray of lumbosacral spine	tendon Unstable knee		

Table 3: Included medication names

Simple analgesics (N02BE*)	Non-steroidal anti-inflammatories (M01A*)	Chondroitin and/or glucosamine (M01AX*)	Topical products for joint and muscular pain (M02A*)	Opioids (N02A*)	Other epileptics (N03AX*)
Caffeine, Paracetamol	Celecoxib	Borate, Chondroitin, Glucosamine, Manganese	Benzydamine	**Weak single opioids**	Gabapentin
Paracetamol	Diclofenac	Borate, Chondroitin, Glucosamine, Manganese	Benzydamine hydrochloride	Codeine	Pregabalin
	Diclofenac potassium	Chrondroitin, Copper, Glucosamine, Manganese, Zinc Sulfate	Cajuput oil, Camphor, Capsicum, Eucalyptus oil, Hydroxybenzoate, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil	Codeine phosphate	
Paracetamol combinations	Diclofenac sodium	Chondroitin, Dimethyl Sulfone, Glucosamine	Cajuput oil, Camphor, Clove, Menthol (TIGER BALM)	Codeine phosphate hemihydrate	
Ibuprofen, Paracetamol	Diclofenac sodium, Misoprostol	Glucosamine	Camphor, Menthol, Eucalyptus oil, Methyl salicylate	Dextropropoxyphene	
	Diclofenac, Misoprostol	Glucosamine, Calcium, Vitamin D, Minerals	Camphor, Eucalyptus oil, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil	Dextropropoxyphene napsylate	
	Etoricoxib	Glucosamine, Chondroitin	Camphor, Menthol, Methyl salicylate	Tramadol	
	Flurbiprofen	Glucosamine hydrochloride	Camphor, Eucalyptus oil, Menthol, Methyl salicylate	Tramadol hydrochloride	
	Ibuprofen	Glucosamine hydrochloride, Chondroitin sulphate	Camphor, Eucalyptus oil, Methyl salicylate, Menthol, Alisma plantago aquatica Root oil extract, Bambusa root	**Combination weak opioid**	
	Ibuprofen lysine	Glucosamine hydrochloride, Chondroitin sulfate, Dimethyl sulfone	Capsaicin	Aspirin, Codeine phosphate	
	Indomethacin	Glucosamine hydrochloride; Chondroitin sulfate, Manganese gluconate, Calcium ascorbate	Capsicum oleoresin, Arnica montana, Arctium lappa root dry, Aloe barbadensis inner leaf juice	Codeine, Ibuprofen	
	Ketoprofen		Diclofenac	Codeine phosphate, Ibuprofen	
	Ketorolac			Codeine, Paracetamol	
	Ketorolac trometamol			Codeine Phosphate, Paracetamol	
	Lumiracoxib			Codeine phosphate hemihydrate, Ibuprofen	
	Mefenamic acid			Dextropropoxyphene, Paracetamol	
	Meloxican				
	Naproxen				
	Naproxen sodium				
	Naproxen, Esomeprazole				

	Parecoxib Parecoxib sodium Piroxicam Rofecoxib Sulindac Tiaprofenic acid	Glucosamine hydrochloride, Calcium, Vitamin D, Vitamin K, Boron Glucosamine hydrochloride; Glucosamine sulfate, Glycine, fructose, Bioflavonoids, Ascorbic acid, Histidine, Lysine hydrochloride, Leucine, Valine, Perna caniculata powder, Calcium pantothenate, Zinc amino acid chelate, Manganese amino acid chelate, Copper gluconate, Selenomethionine Glucosamine, Omega-3 triglycerides Glucosamine sulfate Glucosamine sulfate, Chondroitin sulfate (Shark) Glucosamine sulfate, Shark cartilage Glucosamine sulfate, Potassium chloride Glucosamine sulfate sodium chloride, Eicosapentaenoic acid, Docosahexaenoic acid Ascorbate, Glucosamine, Manganese, Turmeric Borate, Glucosamine, Manganese, Selenium	Diclofenac diethylamine Diclofenac diethylammonium Diclofenac Sodium Ethyl salicylate, Hydroxyethyl salicylate, Methyl salicylate, Nicotinic acid Eucalyptus oil Eucalyptus oil, Pine oil Pumilio, Peppermint oil, Camphor, Methyl salicylate, Menthol, Turpentine oil Eucalyptus oil, Menthol, Methyl salicylate Flurbiprofen sodium Ibuprofen Ketoprofen Menthol Menthol, Camphor, Cajuput oil, Clove oil, Dementholised mint oil Menthol, Camphor, Cajuput oil, Dementholised mint oil, Clove bud oil Menthol, Glycol salicylate Menthol, Eucalyptus oil, Methyl salicylate Methyl salicylate Methyl salicylate, Ethyl salicylate, 2-Hydroxyethyl salicylate, Methyl nicotinate	Dextropropoxyphene napsylate, Paracetamol Tramadol, Paracetamol Tramadol hydrochloride, Paracetamol **Strong single opioids** Fentanyl Fentanyl citrate Hydromorphone Hydromorphone hydrochloride Morphine Morphine hydrochloride Morphine hydrochloride trihydrate Morphine sulfate Morphine sulfate Bp Morphine sulfate pentahydrate Morphine tartrate Oxycodone Oxycodone, Naloxone Oxycodone hydrochloride Oxycodone pectinate Oxycodone hydrochloride, Naloxone hydrochloride	

		Ascorbate, Cod-liver oil, Colecalciferol, Copper, Cyanocobalamin, Folate, Glucosamine, Manganese, Omega-3 triglycerides, Selenium, Tocopherol, Zinc	Methyl salicylate, Eucalyptus oil, Menthol liquid Methyl salicylate, Menthol Nicoboxil/Nonivamide Nonivamide, Butoxyethyl nicotinate Piroxicam Triethanolamine salicylate Trolamine salicylate	Tapentadol Tapentadol hydrochloride	

* Anatomic and Therapeutic Classifications (ATC) category

Table 4: Test name subgroups for low back, knee, shoulder, and neck imaging tests and procedures

Low back imaging subgroups	Knee imaging subgroups	Shoulder imaging subgroups	Neck imaging subgroups
Lumbosacral plain radiograph*	Knee plain radiograph*	Shoulder plain radiograph*	Neck plain radiograph*
Lumbosacral CT*	Knee CT*	Shoulder ultrasound*	Neck CT*
Lumbosacral MRI*	Knee MRI*	Shoulder MRI*	Neck MRI*
Lumbosacral injection*	Knee injection*	Shoulder injection*	Neck injection*
Lumbosacral unspecified*#	Knee unspecified*#	Shoulder unspecified*#	Neck unspecified*#
Lumbosacral ultrasound^	Knee ultrasound*	Shoulder hydrodilatation*	Neck ultrasound^
Lumbosacral other^	Knee other^	Shoulder other^	Neck other^
	Knee aspiration^	Shoulder aspiration^	Neck aspiration^
	Knee arthrogram^	Shoulder arthrogram^	
		Shoulder CT^	
		Shoulder fluoroscopy^	

* Eligible

^ Ineligible

Analyse as plain radiograph

AUTHOR CONTRIBUTIONS

RH, DOC and RB conceived the study. LB and AG were responsible for data coding and the statistical analysis plan. CP provided expertise in the use of the POLAR database. DM provided clinical context in managing musculoskeletal conditions in the general practice setting. All authors contributed to refining the protocol and approved the submitted protocol.

FUNDING STATEMENT

This work was supported by an Arthritis Queensland, Arthritis South Australia and the Allan and Beryl Stephens Grant from Arthritis Australia (ID N/A). Arthritis Australia did not contribute to the conduct of this study. It is also supported by an Australian National Health and Medical Research Council (NHMRC) Program Grant (APP1113532). DOC is supported by a TRIP Fellowship and RB is supported by an NHMRC Investigator Fellowship.

COMPETING INTERESTS STATEMENT

RH, DOC, RB and DM report grants from Arthritis Australia (not-for-profit organisation), during the conduct of the study. CP is an employee of Outcome Health, the not-for-profit organisation that developed the POLAR database and chairs the Product improvement group of the Australian Digital Health Agency. It has no relationship with the research but has provided grant funding to Outcome Health. LB reports consultancy fees paid to Monash University from Charite Medical University Berlin, Jesuit Social Services Victoria, and Swinburne University of Technology, outside the submitted work.

REFERENCES

1. Royal Australian College of General Practitioners. General Practice: Health of the Nation. East Melbourne, Vic: RACGP, 2019.
2. Britt H MG, Henderson J, Bayram C, Harrison C, Valenti L, Pan Y, Charles J, Pollack AJ, Wong C, Gordon J. General practice activity in Australia 2015–16. General practice series no. 40. Sydney: Sydney University Press, 2016.
3. Rahman N, Gruber D. The burden of musculoskeletal conditions in Australia: a detailed analysis of the Australian Burden of Disease Study 2011: Australian Institute of Health and Welfare 2017.
4. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020;396(10258):1204-22.
5. Britt H MG, Bayram C, Henderson J, Valenti L, Harrison C, Pan Y, Charles J,, Pollack AJ CT, Gordon J, Wong C. A decade of Australian general practice activity 2006–07 to 2015–16. General practice series no. 41. Sydney: Sydney University Press, 2016.
6. Bennell KL BC, Harrison C, Brand C, Buchbinder R, Haas R, and Hinman RS. Trends in management of hip and knee osteoarthritis in general practice in Australia over an 11-year window: a nationwide cross-sectional survey. *The Lancet Regional Health – Western Pacific* 2021;12 doi: https://authors.elsevier.com/sd/article/S2666606521000961
7. Naunton J, Harrison C, Britt H, et al. General practice management of rotator cuff related shoulder pain: A reliance on ultrasound and injection guided care. *Plos one* 2020;15(1):e0227688.
8. Busingye D, Gianacas C, Pollack A, et al. Data Resource Profile: MedicineInsight, an Australian national primary health care database. *International Journal of Epidemiology* 2019;48(6):1741-41h.
9. Pearce C, McLeod A, Rinehart N, et al. What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program. *Stud Health Technol Inform* 2019;264:303-07.
10. de Lusignan S, Sun B, Pearce C, et al. Coding errors in an analysis of the impact of pay-for-performance on the care for long-term cardiovascular disease: a case study. *Journal of Innovation in Health Informatics* 2014;21(2):92-101.
11. Hemkens LG, Benchimol EI, Langan SM, et al. The reporting of studies using routinely collected health data was often insufficient. *Journal of clinical epidemiology* 2016;79:104-11.
12. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885.
13. Healthmap: Healthdirect Australia; [Available from: https://studio.healthmap.com.au/#/h4f accessed 27th September 2020.
14. The State of Victoria Department of Premier and Cabinet. Population Diversity in Victoria: 2016 Census Local Government Areas, 2018.
15. Yan J, Hawes L, Turner L, et al. Antimicrobial prescribing for children in primary care. *Journal of paediatrics and child health* 2019;55(1):54-58.
16. Turner LR, Pearce C, Borg M, et al. Characteristics of patients presenting to an after-hours clinic: results of a MAGNET analysis. *Australian journal of primary health* 2017;23(3):294-99.
17. Turner LR, Cicuttini F, Pearce C, et al. Cardiovascular disease screening in general practice: General practitioner recording of common risk factors. *Preventive medicine* 2017;99:282-85.
18. Sezgin G, Georgiou A, Hardie R-A, et al. Compliance with pathology testing guidelines in Australian general practice: protocol for a secondary analysis of electronic health record data. *BMJ open* 2018;8(11)
19. Australian Institute of Health and Welfare. Musculoskeletal conditions and comorbidity in Australia. Canberra: AIHW, 2019.

20. Pearce C, McLeod A, Patrick J, et al. Coding and classifying GP data: the POLAR project. *BMJ health & care informatics* 2019;26(1)
21. Cotd R, Rothwell D, Palotay J, et al. The Systematized Nomenclature of Human and Veterinary Medicine-SNOMED International. *Northfield, IL: College of American Pathologists* 1993
22. World Health Organization. WHO collaborating Centre for Drug Statistics Methodology: ATC classification index with DDDs and guidelines for ATC classification and DDD assignment 2020. *Oslo, Norway: Norwegian Institute of Public Health* 2019
23. Hovis KK, Alizai H, Tham S-C, et al. Non-traumatic anterior cruciate ligament abnormalities and their relationship to osteoarthritis using morphological grading and cartilage T2 relaxation times: data from the Osteoarthritis Initiative (OAI). *Skeletal radiology* 2012;41(11):1435-43.
24. Buchbinder R, Underwood M, Hartvigsen J, et al. The Lancet Series call to action to reduce low value care for low back pain: an update. *Pain* 2020;161(1):S57.
25. Mathieson S, Wertheimer G, Maher C, et al. What proportion of patients with chronic noncancer pain are prescribed an opioid medicine? Systematic review and meta-regression of observational studies. *Journal of Internal Medicine* 2020;287(5):458-74.
26. Ittaman SV, VanWormer JJ, Rezkalla SH. The role of aspirin in the prevention of cardiovascular disease. *Clinical medicine & research* 2014;12(3-4):147-54.
27. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology* 2015;44(3):827-36.
28. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PloS one* 2014;9(6):e99825.
29. Australian Bureau of Statistics. Socio-economic indexes for areas (SEIFA) - technical paper 2016. *cat no 20330 55001* 2018
30. Halpin B. SADI: Sequence analysis tools for Stata. *The Stata Journal* 2017;17(3):546-72.
31. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *arXiv preprint arXiv:200300381* 2020
32. Imai C, Hardie R-A, Franco GS, et al. Harnessing the potential of electronic general practice pathology data in Australia: An examination of the quality use of pathology for type 2 diabetes patients. *International Journal of Medical Informatics* 2020;141:104189.
33. Harrison C, Britt H, Miller G, et al. Prevalence of chronic conditions in Australia. *PloS one* 2013;8(7):e67494.

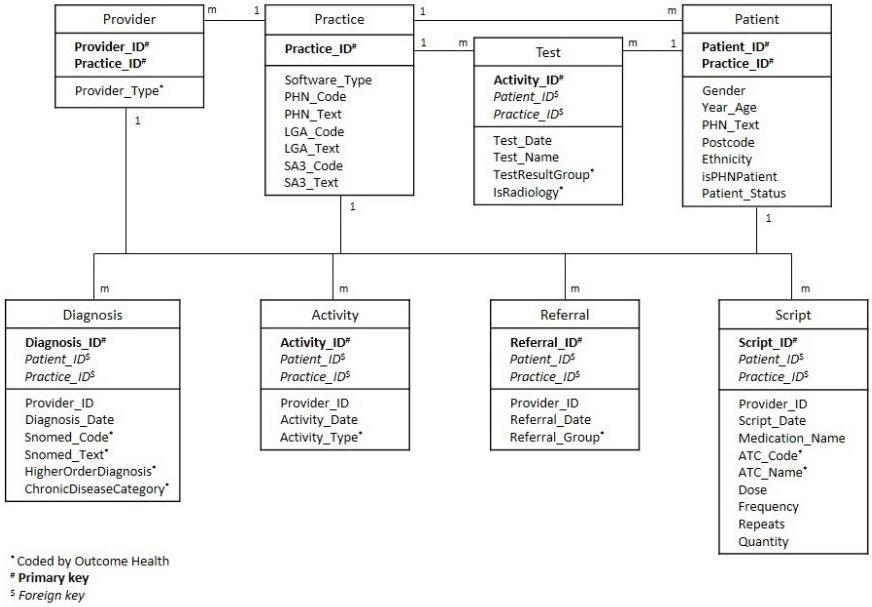

Figure 1

271x186mm (96 x 96 DPI)

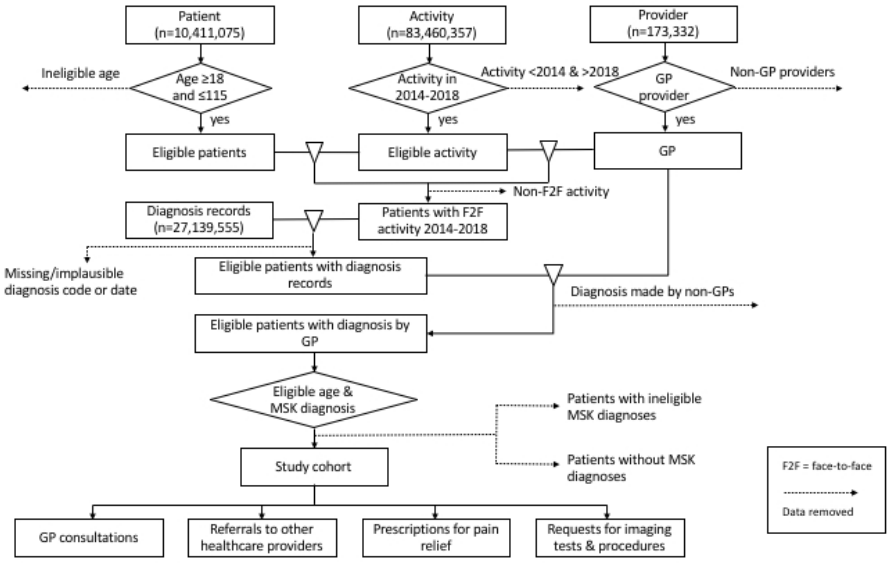

Figure 2

60x34mm (300 x 300 DPI)

Appendix 1. Initial string match terms used to code imaging records

Imaging region strings	X-ray strings	MRI strings	CT strings	Ultrasound strings
Knee				
KNEE	X-RAY KN	MRI KN	CT KN	US KN
KN	X-RAY RIGHT KN	MRI RIGHT KN	CT RIGHT KN	US RIGHT KN
BOTH K*	X-RAY LEFT KN	MRI LEFT KN	CT LEFT KN	US LEFT KN
	XRAY KN	MR KN	KNEE CT	ULTRASOUND KN
Exclude:	XRAY RIGHT KN	MR RIGHT KN		ULTRASOUND RIGHT KN
KNOW	XRAY LEFT KN	MR LEFT KN		ULTRASOUND LEFT KN
KNIGHT	XR KN	MAGNETIC RESONANCE KN		KNEE US
	XR RIGHT KN	MAGNETIC RESONANCE RIGHT KN		KNEE ULT
	XR LEFT KN	MAGNETIC RESONANCE LEFT KN		
	PLAIN FILM KN	KNEE MR		
	PLAIN FILM RIGHT KN	KNEE MAGNETIC		
	PLAIN FILM LEFT KN			
	RADIOGRAPH KN			
	RADIOGRAPH RIGHT KN			
	RADIOGRAPH LEFT KN			
	KNEE X			
	KNEE RADIOGR			
	KNEE PLAIN FILM			
Shoulder				
SHOULDER	X-RAY SH	MRI SH	CT SH	US SH
SH	X-RAY RIGHT SH	MRI RIGHT SH	CT RIGHT SH	US RIGHT SH
CLAVICLE*	X-RAY LEFT SH	MRI LEFT SH	CT LEFT SH	US LEFT SH
	XRAY SH	MR SH	SHOULDER CT	ULTRASOUND SH
Exclude:	XRAY RIGHT SH	MR RIGHT SH		ULTRASOUND RIGHT SH
SHBG	XRAY LEFT SH	MR LEFT SH		ULTRASOUND LEFT SH
TSH	XR SH	MAGNETIC RESONANCE SH		SHOULDER US
SHEET	XR RIGHT SH	MAGNETIC RESONANCE RIGHT SH		SHOULDER ULT
FSH	XR LEFT SH	MAGNETIC RESONANCE LEFT SH		
GSHS	PLAIN FILM SH	SHOULDER MR		

PSH SH/	PLAIN FILM RIGHT SH PLAIN FILM LEFT SH RADIOGRAPH SH RADIOGRAPH RIGHT SH RADIOGRAPH LEFT SH SHOULDER X SHOULDER RADIOGR SHOULDER PLAIN FILM	SHOULDER MAGNETIC		
Neck				
NECK NEC CERVIC C1 C2 C3 C4 C5 C6 C7 C SPINE SPINE CX* **Exclude:** FEMORAL NECK CERVICAL CYTOLOGY	X-RAY NECK XRAY NECK XR NECK X-RAY CERVICAL XRAY CERVICAL XR CERVICAL PLAIN FILM NECK PLAIN FILM CERVICAL RADIOGRAPH NECK RADIOGRAPH CERVICAL NECK X CERVICAL X NECK PLAIN FILM CERVICAL PLAIN FILM NECK RADIOGRAPH CERVICAL RADIOGRAPH	MRI NECK MR NECK MRI CERVICAL MR CERVICAL NECK MR CERVICAL MR	CT NECK CT CERVICAL NECK CT CERVICAL CT	US NECK US CERVICAL ULTRASOUND NECK ULTRASOUND CERVICAL NECK US CERVICAL US NECK ULT CERVICAL ULT
Low back				
LUMB SACR L1 L2 L3 L4 L5	X-RAY LUMB XRAY LUMB XR LUMB X-RAY SACR XRAY SACR XR SACR PLAIN FILM LUMB	MRI LUMB MR LUMB MRI SACR MR SACR LUMBAR MR SACRAL MR	CT LUMB CT SACR LUMBAR CT SACRAL CT	US LUMB ULTRASOUND LUMB US SACR ULTRASOUND SACR LUMBAR US LUMBAR ULT SACRAL US

LOWER BACK **Exclude:** FUNGAL2	PLAIN FILM SACR RADIOGRAPH LUMB RADIOGRAPH SACR LUMBAR X LUMBAR RADIOGRAPH LUMBAR PLAIN FILM SACRAL X SACRAL RADIOGRAPH SACRAL PLAIN FILM			SACRAL ULT

*String match term added after initial coding

Appendix 2

The RECORD statement – checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	a) Study design (observational cohort study) is included in the title b) Abstract (methods and analysis) contains a summary of what was done. As this is a protocol, what was found is not applicable	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title & abstract Methods and analysis of Abstract refers to geographic region (South Eastern Victoria) and timeframe within the study (1/1/14 to 31/12/18) N/A
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction contains rationale for protocol (explicit reporting of the systematic approach used to classify and select eligible records from the POLAR database will facilitate replication and transparency)		
Objectives	3	State specific objectives, including any prespecified hypotheses	Objectives No prespecified hypotheses reported as this is a protocol		
Methods					
Study Design	4	Present key elements of study design early in the paper	Study design section of Methods (retrospective cohort study)		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Setting section of Methods includes locations of practices with POLAR database, dates of the study period (exposure and data collection) and follow-up		

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	Eligibility criteria are presented in Table 2. Sources and methods of selection are described in Variables section of Methods. Setting – patient-level follow-up data until 31/12/18 will be included N/A	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Variables section of Methods Diagnoses and imaging records within Variables section of Methods N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Variables section of Methods and Tables 3, 4 & 5	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Variables section of Methods and Tables 3, 4 & 5
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Data source section of Methods		
Bias	9	Describe any efforts to address potential sources of bias	Variables and data cleaning sections of Methods		
Study size	10	Explain how the study size was arrived at	Sample size consideration		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Analyses section of Methods		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	a) Analyses section of Methods b) N/A		

		(b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	c) Approach to dataset creation d) Data access and cleaning – only data after which there has been consistent reporting within a practice will be included e) Analyses – sensitivity analysis to include entire follow-up period (instead of 1 year) for each participant and based on date of most recent diagnosis instead of index diagnosis for participants with multiple body regions affected by an eligible musculoskeletal diagnosis		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Data access and cleaning methods Data access and cleaning methods
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	N/A
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	N/A	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	N/A
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on	N/A		

		exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	N/A		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	N/A		
Discussion					
Key results	18	Summarise key results with reference to study objectives	N/A		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	N/A		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion – potential representativeness of the POLAR database and generalizability to the wider population is discussed		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Funding statement		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Appendix 1 – initial string match terms used to code imaging tests

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.