








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

Romi Haas <sup>1,2</sup>, Ljoudmila Busija <sup>1</sup>, Alexandra Gorelik <sup>1,2</sup>,  
Denise A O'Connor <sup>1,2</sup>, Christopher Pearce <sup>3</sup>, Danielle Mazza <sup>4</sup>,  
Rachelle Buchbinder <sup>1,2</sup>

**To cite:** Haas R, Busija L, Gorelik A, *et al*. Patterns of care for people presenting to Australian general practice with musculoskeletal complaints based on routinely collected data: protocol for an observational cohort study using the Population Level Analysis and Reporting (POLAR) database. *BMJ Open* 2021;**11**:e055528. doi:10.1136/bmjopen-2021-055528

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-055528>).

Received 16 July 2021  
Accepted 06 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Romi Haas;  
[romi.haas@monash.edu](mailto:romi.haas@monash.edu)

## 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 southeastern Victoria. Data for patients with a low back, neck, shoulder and/or knee condition and who received at least one general practitioner (GP) face-to-face consultation between 1 January 2014 and 31 December 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 December 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

## 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 POpulation Level Analysis and Reporting (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 general practitioner (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.

and 16975, respectively). Study findings will be reported to Outcome Health, participating PHNs, disseminated in academic journals and presented in conferences.

## INTRODUCTION

General practice plays an essential role in providing primary healthcare to the population. In Australia 86% of the population visits a general practitioner (GP) multiple times

a year,<sup>1</sup> and nearly 20% of these consultations are for a musculoskeletal condition.<sup>2</sup> These conditions account for 23% of the years lived with disability in Australia<sup>3</sup> and are also a major cause of disability worldwide.<sup>4</sup> Until 2016, the BEACH (Bettering the Evaluation and Care of Health) programme provided the most comprehensive data on clinical activities of Australian general practice.<sup>5</sup> The programme 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.<sup>6,7</sup> However, in-depth exploration of these activities within the BEACH programme 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 data sets 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<sup>8</sup> and the POpulation Level Analysis and Reporting (POLAR) databases<sup>9</sup> are examples of longitudinal general practice data sets within Australia. Unlike POLAR, the Medicine Insight programme does not currently include referrals provided by GPs to other healthcare providers.<sup>8</sup> 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 healthcare, 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.<sup>10</sup> 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.<sup>11</sup> 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 9 (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.<sup>12</sup>

The POLAR database draws data from every consultation occurring for millions of patients in approximately 30% of general practices across southeastern Victoria,<sup>13</sup> an area

that comprises more than half of Victoria's population.<sup>14</sup> 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.

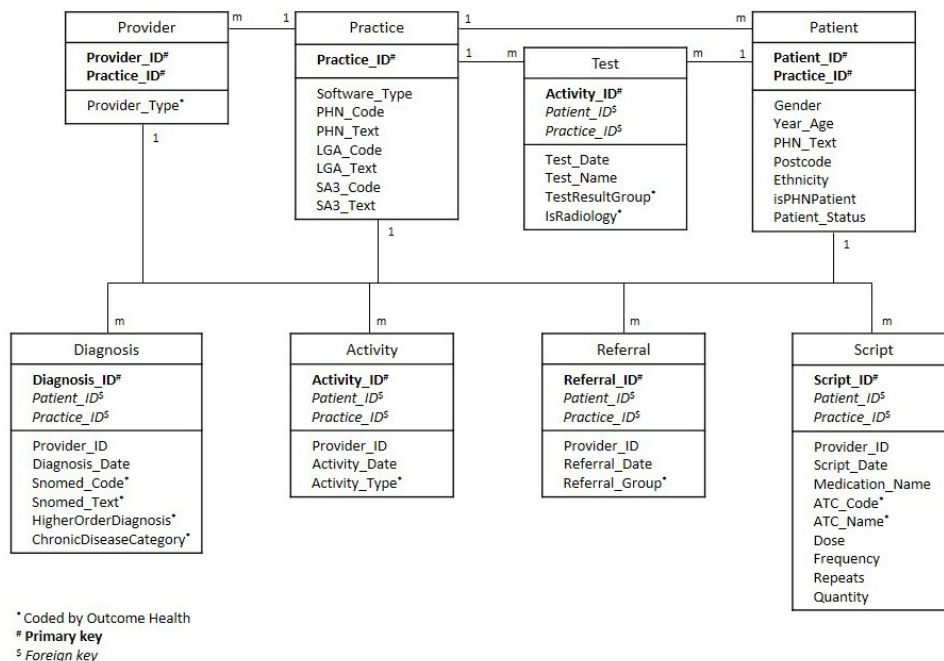
Previous studies have used the POLAR database to investigate patterns of antimicrobial prescribing for children,<sup>15</sup> to examine characteristics of patients presenting to an after-hours clinic,<sup>16</sup> to estimate GP recording of cardiovascular risk factors<sup>17</sup> and to describe characteristics of pathology test ordering in general practice.<sup>18</sup> 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-related and practice-related variables.
4. Examine the longitudinal changes in GP management for these conditions between 2014 and 2018 inclusive.



**Figure 1** Database structure. ATC, Anatomical Therapeutic Chemical; PHN, primary health networks.

## 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.<sup>9</sup> 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 are extracted from two different clinical information systems, covering 90% of included general practices. All data are extracted using the Hummingbird data extraction tool.<sup>9</sup>

### Setting

The POLAR database contains de-identified patient-related data from all electronic medical records of consenting general practices within the primary health networks (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 31 December 2018. Data analyses will be completed by the end of 2021.

### Participants

The study cohort will include people diagnosed during 2014–2018 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.<sup>19</sup> Eligibility criteria are presented in table 1. We excluded traumatic diagnoses and conditions typically primarily managed by a specialist (eg, 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 an eligible diagnosis could result from consultation with other healthcare 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 (eg, diagnosis records), we used this coding to select eligible records that fitted our inclusion criteria. In circumstances where there was no coding (eg, 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 followed by a manual process for quality checking and correction.<sup>20</sup> 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<sup>21</sup> and prescriptions are coded according to the Anatomical Therapeutic Chemical (ATC) classification system.<sup>22</sup> In cases where

**Table 1** Eligibility criteria

Patient population			Patient management				
Diagnoses	Provider	Patient	Practice	Activity	Referrals	Prescriptions	Imaging tests and procedures
Low back Knee Shoulder Neck	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, eg, 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 hydrodilataion Cervical plain radiograph Cervical CT Cervical MRI Cervical injection
<b>Exclude:</b> Trauma Systemic inflammatory arthritis							

CT, Computed Tomography; MRI, Magnetic Resonance Imaging.

there is no standardised classification system available (eg, providers and referrals), Outcome Health used a similar process to code these variables into relevant categories (eg, type of healthcare 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.<sup>20</sup> 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 (eg, 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 in [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.<sup>23</sup> 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 used key chronic disease groups as a qualifier.<sup>9</sup> For example, free text such as ‘low back pain’ or ‘angina’ could be qualified as a chronic disease if present for 6 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.



**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 arthrodiesis
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	Inflammation of rotator cuff tendon	CT of cervical spine
CT of spine	Complete tear, knee, medial collateral ligament	Impingement syndrome of shoulder region	CT of neck
Curvature of spine	Contusion of knee	Inflammation of rotator cuff tendon	Degeneration of cervical intervertebral disc
Decompression laminectomy	Derangement of knee	Injury of glenoid labrum of shoulder joint	Diffuse cervicobrachial syndrome
Decompression of lumbar spine	Disorder of patellofemoral joint	Injury of shoulder region	Excision of cervical intervertebral disc
Degeneration of intervertebral disc	Finding of tear meniscus	MRI of shoulder	Injury of cervical spine
Degeneration of lumbar intervertebral disc	Fracture of tibial plateau	Osteoarthritis of acromioclavicular joint	Kyphoscoliosis deformity of spine
Diagnostic radiography of coccyx	Haemarthrosis of knee	Osteoarthritis of shoulder	Kyphosis deformity of spine
Discitis	Inflammation of bursa of patella	Painful arc syndrome	MRI of neck
Discogenic pain	Injury of anterior cruciate ligament	Radiography of shoulder	MRI of cervical spine
Disorder of joint of spine	Injury of knee	Repair of musculotendinous cuff of shoulder	Muscle spasm of cervical muscle of neck
Disorder of vertebra	Knee joint—varus deformity	Repair of shoulder	Neck injury
Exploration of spine	Knee joint effusion	Rotator cuff impingement syndrome	Neck pain
Facet joint pain	Knee joint valgus deformity	Rotator cuff syndrome	Neck sprain
Fracture of body of vertebra	Knee locking	Rupture of tendon of biceps	Neck structure
Fracture of lumbar spine	Knee pain	Rupture of tendon of biceps, long head	Pain in cervical spine
Fracture of sacrum	Knee region structure	Shoulder pain	Prolapsed cervical intervertebral disc
Fracture of vertebral column	Knee stiff	Shoulder reconstruction	Radiography of cervical spine
Injury of back	Loose body in knee	Shoulder region structure	Spinal stenosis in cervical region
Injury of coccyx	MRI of knee	Shoulder strain	Stiff neck
Intervertebral disc prolapse	Osteoarthritis of knee	Shoulder tendinitis	Strain of neck muscle
L4/L5 disc	Osteotomy of proximal tibia	Sprain of acromioclavicular ligament	Strain of tendon of neck
L5/S1 disc	Osteotomy of tibia	Sprain of shoulder	Torticollis
Laminectomy	Patellar instability	Structure of left shoulder region	Whiplash injury to neck
Lordosis deformity of spine	Patellar maltracking	Structure of right shoulder region	
Low back pain	Patellar tendonitis	Structure of rotator cuff including muscles and tendons	
Low back strain	Patellectomy	Subacromial bursitis	
Lower back injury	Patellofemoral osteoarthritis	Subdeltoid bursitis	
Lower back structure	Patellofemoral stress syndrome	Subluxation of acromioclavicular joint	
Lumbar	Prepatellar bursitis	Subscapularis tendinitis	
Lumbar discectomy	Problem knee	Supraspinatus tear	
Lumbar laminectomy	Radiological examination of knee	Supraspinatus tendinitis	
Lumbar microdiscectomy	Repair of anterior cruciate ligament of knee joint	Total shoulder replacement	
Lumbar radiculopathy	Repair of knee collateral ligaments	US shoulder region	
Lumbar region back structure	Repair of knee cruciate ligaments		
Lumbar spinal fusion	Repair of meniscus		
Lumbar sprain	Repair of patellar tendon		
Lumbosacral spine	Replacement of total knee joint		
Lumbosacral spondylosis	Rupture of anterior cruciate ligament		
Lumbosacral spondylosis without myelopathy	Rupture of cruciate ligaments		
Lumbosacral strain	Rupture of medial collateral ligament of knee		
Lumbosacral radiculopathy	Rupture of posterior cruciate ligament		
MRI of spine	Sprain of knee		
Manipulation of spine	Sprain of lateral collateral ligament of knee		
MRI of lumbar spine	Sprain of medial collateral ligament of knee		
Nerve root compression syndrome	Stabilisation of patellofemoral joint		
Nerve root disorder	Strain of knee		
Operative procedure on spinal structure	Strain of patellar tendon		
Osteoarthritis of lumbar spine	Strain of tendon of medial thigh muscle		
Pain in lumbar spine	Structure of left knee		
Pain in the coccyx	Structure of prepatellar bursa		
Prolapsed lumbar intervertebral disc	Structure of right knee		
Radiography of spine	Subluxation of patellofemoral joint		
Sacral back pain	Suprapatellar bursitis		
Sacroiliac arthrodesis	Swollen knee		
Sacroiliac joint inflamed	Synovial cyst of knee		
Sacroiliac joint pain	Synovial cyst of popliteal space		
Scoliosis deformity of spine	Tear of lateral meniscus of knee		
Scoliosis of lumbar spine	Tear of medial meniscus of knee		
Spasm of back muscles	Tear of meniscus of knee		
Spinal arthritis deformans	Total knee replacement		
Spinal arthrodesis	Total replacement of left knee joint		
Spinal claudication	Total replacement of right knee joint		
Spinal injury	Traumatic rupture of patellar tendon		
Spinal stenosis	Unstable knee		
Spinal stenosis of lumbar region			
Spondylitis			
Spondylolisthesis			
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			

CT, Computed Tomography; MRI, Magnetic Resonance Imaging; US, Ultrasound.

### Activity records

Activity records are coded in POLAR according to the type of consultation provided (eg, Telehealth, visit, telephone). Each time a note is recorded in the narrative

section it is coded by the electronic medical record (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 (ie, 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 (eg, 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 ATC system.<sup>22</sup> 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 such as paracetamol; non-steroidal anti-inflammatories (NSAIDs); chondroitin and/or glucosamine; topical products for joint and muscular pain; opioids; gabapentinoids and any relevant combinations. We included gabapentinoids such as gabapentin and pregabalin because these are being increasingly used for the management of musculoskeletal conditions such as non-specific low back pain or sciatica despite evidence of a lack of effectiveness and a higher risk of adverse events.<sup>24</sup> Opioid analgesics were further categorised into (i) weak single ingredient opioid analgesics (eg, codeine), defined as <50 morphine milligram equivalents (MME) per day; (ii) strong single ingredient opioid analgesics (eg, tapentadol, oxycodone, morphine), defined as 150 MME per day; and (iii) combination opioid analgesics.<sup>25</sup> 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 (eg, pseudoephedrine), antihistamines (eg, doxylamine), opioid cough suppressants (eg, dextromethorphan) and expectorants (eg, guaifenesin). These were excluded on the basis that they were likely to have been prescribed for another condition (eg, aspirin for secondary prevention of cardiovascular disease).<sup>26</sup>

### 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 (ie, 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 (eg, 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 online supplemental appendix 1.

During the coding process, there were numerous test names that did not definitively identify a type of imaging test (eg, ‘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 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 96.0%, 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 UK.<sup>27</sup>

**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*)	Gabapentinoids (N03AX*)
[Caffeine, Paracetamol] Paracetamol	Celecoxib Diclofenac Diclofenac potassium Diclofenac sodium [Diclofenac sodium, Misoprostol] Misoprostol [Diclofenac, Misoprostol] Etoricoxib Flurbiprofen Ibuprofen Ibuprofen lysine Indomethacin Ketoprofen Ketorolac Ketorolac trometamol Lumiracoxib Mefenamic acid Meloxicam Naproxen Naproxen sodium [Naproxen, Esomeprazole] Parecoxib Parecoxib sodium Piroxicam Rofecoxib Sulindac Tiaprofenic acid	[Borate, Chondroitin, Glucosamine, Manganese] [Chondroitin, Copper, Glucosamine, Manganese, Zinc Sulfate] [Chondroitin, Dimethyl Sulfone, Glucosamine] Glucosamine [Glucosamine, Calcium, Vitamin D, Minerals] [Glucosamine, Chondroitin] Glucosamine hydrochloride [Glucosamine hydrochloride, Chondroitin sulfate] [Glucosamine hydrochloride, Chondroitin sulfate, Dimethyl sulfone] [Glucosamine hydrochloride, Chondroitin sulfate, Manganese gluconate, Calcium ascorbate] [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] [Ascorbate, Cod-liver oil, Colecalciferol, Copper, Cyanocobalamin, Folate, Glucosamine, Manganese, Omega-3 triglycerides, Selenium, Tocopherol, Zinc]	Benzydamine Benzydamine hydrochloride [Cajuput oil, Camphor, Capsicum, Eucalyptus oil, Hydroxybenzoate, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil] [Cajuput oil, Camphor, Clove, Menthol (Tiger Balm)] [Camphor, Menthol, Eucalyptus oil, Methyl salicylate] [Camphor, Eucalyptus oil, Mentha X Piperita, Menthol, Methyl salicylate, Pinus, Turpentine oil] [Camphor, Menthol, Methyl salicylate] [Camphor, Eucalyptus oil, Menthol, Methyl salicylate] [Camphor, Eucalyptus oil, Methyl salicylate, Menthol, Alisma plantago aquatica Root oil extract, Bambusa root] Capsaicin [Capsicum oleoresin, Arnica montana, Arctium lappa root dry, Aloe barbadensis inner leaf juice] Diclofenac 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] [Methyl salicylate, Eucalyptus oil, Menthol liquid] [Methyl salicylate, Menthol] Nicoboxil/Nonivamide [Nonivamide, Butoxyethyl nicotinate] Piroxicam Triethanolamine salicylate Trolamine salicylate	<b>Weak single opioids</b> Codeine Codeine phosphate Codeine phosphate hemihydrate Dextropropoxyphene Dextropropoxyphene napsylate Tramadol Tramadol hydrochloride  <b>Combination weak opioid</b> [Aspirin, Codeine phosphate] [Codeine, Ibuprofen] [Codeine phosphate, Ibuprofen] [Codeine, Paracetamol] [Codeine phosphate hemihydrate, Ibuprofen] [Dextropropoxyphene, Paracetamol] [Dextropropoxyphene napsylate, Paracetamol] [Tramadol, Paracetamol] [Tramadol hydrochloride, Paracetamol]  <b>Strong single opioids</b> 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] Tapentadol Tapentadol hydrochloride	Gabapentin Pregabalin

\*Anatomic and Therapeutic Classifications (ATC) category.

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 1 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

**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 hydrodilataion*	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.

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 data set creation

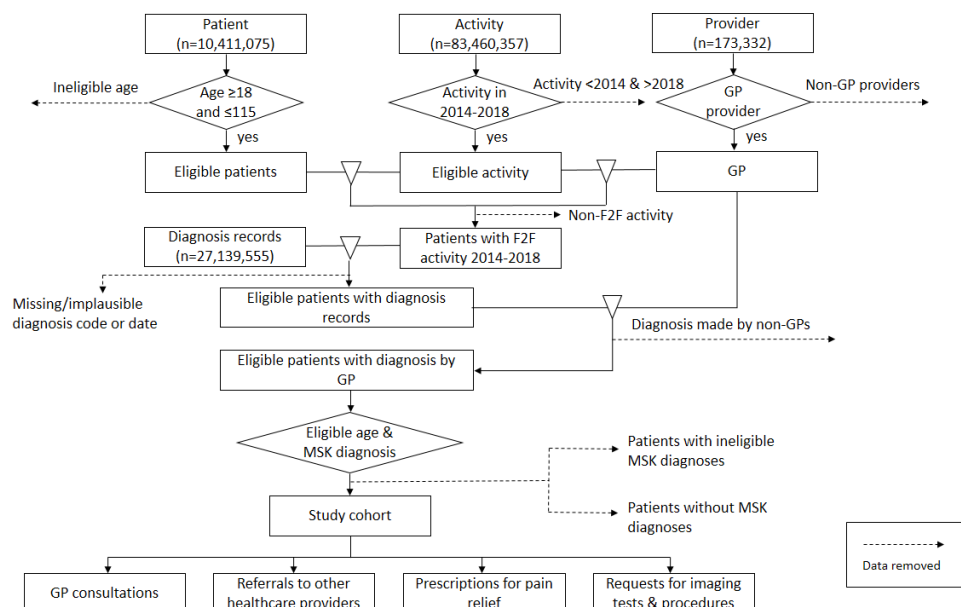
We will use a systematic process to exclude ineligible records in order to merge data and select the study cohort (figure 2). This process will require the merging of five relational data files (patient, practice, provider, activity and diagnosis) in a specific sequence to ensure all relevant records 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 V.15 (StataCorp LP) for data management and analyses. The methods in this protocol are structured according to RECORD guidelines (online supplemental appendix 2).<sup>12</sup> For variables with a recognised coding system, full lists of codes used to define eligible variables are available from <https://clinicalcodes.rss.mhs.man.ac.uk/medcodes/article/174/>.<sup>28</sup>

Descriptive statistics will be used to summarise the study cohort including the number and type of eligible musculoskeletal conditions, patient demographics and

**Figure 2** Approach to data set creation. GP, general practitioner; MSK, musculoskeletal; F2F, face-to-face.



comorbidities. These will be compared with 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 31 December 2018. For prescriptions, the primary analysis will include the entire follow-up period because repeated prescriptions over more than 1 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 (eg, 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 gabapentinoids. 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 5-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. 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-related and practice-related characteristics will be examined using regression analysis. Predictors will include patient gender, socioeconomic status, residential location, body region(s) affected by eligible musculoskeletal conditions and PHN of the practice with adjustment for age and time since index diagnosis. Socioeconomic status will be defined by the Index of Relative Socioeconomic Advantage and Disadvantage using 2016 Census data.<sup>29</sup>

Sequence analysis will be used to categorise sequences of management types of people with eligible musculoskeletal conditions into similar groups based on observed characteristics.<sup>30</sup> 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-related 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 20–30 subjects per subgroup,<sup>31</sup> 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.<sup>32</sup> 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,<sup>33</sup> 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, neck, shoulder and knee 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.<sup>32</sup> 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.<sup>20</sup> 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, imaging requests and prescriptions are not mandated in the source EMRs. 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 were 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.

### Author affiliations

<sup>1</sup>Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

<sup>2</sup>Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Health, Malvern, Victoria, Australia

<sup>3</sup>Outcome Health, Blackburn, Victoria, Australia

<sup>4</sup>Department of General Practice, Monash University, Notting Hill, Victoria, Australia

**Contributors** RH, DAO'C 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** 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) Programme Grant (APP1113532). DAO'C is supported by a TRIP Fellowship and RB is supported by an NHMRC Investigator Fellowship (APP1194483).

**Competing interests** RH, DAO'C, 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.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iDs

Romi Haas <http://orcid.org/0000-0001-9100-5509>

Ljoudmila Busija <http://orcid.org/0000-0001-7464-9089>

Alexandra Gorelik <http://orcid.org/0000-0003-1032-0457>

Denise A O'Connor <http://orcid.org/0000-0002-6836-122X>

Christopher Pearce <http://orcid.org/0000-0001-5371-8196>

Danielle Mazza <http://orcid.org/0000-0001-6158-7376>

Rachelle Buchbinder <http://orcid.org/0000-0002-0597-0933>

## REFERENCES

- 1 Royal Australian College of General Practitioners. *General practice: health of the nation*. East Melbourne, Vic: RACGP, 2019.
- 2 Britt HMG, Henderson J, Bayram C. *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. *Lancet* 2020;396:1204–22.
- 5 Britt HMG, Bayram C, Henderson J. *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. 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 <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:e0227688.
- 8 Busingye D, Gianacas C, Pollack A, et al. Data resource profile: MedicineInsight, an Australian National primary health care database. *Int J Epidemiol* 2019;48:1741–41h.
- 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;264:303–7.
- 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. *J Innov Health Inform* 2014;21:92–101.
- 11 Hemkens LG, Benchimol EI, Langan SM, *et al*. The reporting of studies using routinely collected health data was often insufficient. *J Clin Epidemiol* 2016;79:104–11.
  - 12 Benchimol EI, Smeeth L, Guttman A, *et al*. The reporting of studies conducted using observational Routinely-collected health data (record) statement. *PLoS Med* 2015;12:e1001885.
  - 13 Healthmap: Healthdirect Australia. Available: <https://studio.healthmap.com.au/#/h4f> [Accessed 27 Sep 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. *J Paediatr Child Health* 2019;55:54–8.
  - 16 Turner LR, Pearce C, Borg M, *et al*. Characteristics of patients presenting to an after-hours clinic: results of a magnet analysis. *Aust J Prim Health* 2017;23:294–9.
  - 17 Turner LR, Cicuttini F, Pearce C, *et al*. Cardiovascular disease screening in general practice: general practitioner recording of common risk factors. *Prev Med* 2017;99:282–5.
  - 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:e024223.
  - 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 Inform* 2019;26:e100009.
  - 21 Cote R, Rothwell D, Palotay J. *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 Radiol* 2012;41:1435–43.
  - 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 Suppl 1:S57–64.
  - 25 Mathieson S, Wertheimer G, Maher CG, *et al*. What proportion of patients with chronic noncancer pain are prescribed an opioid medicine? systematic review and meta-regression of observational studies. *J Intern Med* 2020;287:458–74.
  - 26 Ittaman SV, VanWormer JJ, Rezkalla SH. The role of aspirin in the prevention of cardiovascular disease. *Clin Med Res* 2014;12:147–54.
  - 27 Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data resource profile: clinical practice research Datalink (CPRD). *Int J Epidemiol* 2015;44:827–36.
  - 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: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. *Stata J* 2017;17:546–72.
  - 31 Dalmajer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *arXiv preprint arXiv* 2020:200300381.
  - 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. *Int J Med Inform* 2020;141:104189.
  - 33 Harrison C, Britt H, Miller G, *et al*. Prevalence of chronic conditions in Australia. *PLoS One* 2013;8:e67494.