Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases

Kelvin Jordan, Alexandra M Clarke, Deborah PM Symmons, Douglas Fleming, Mark Porcheret, Umesh T Kadam and Peter Croft

ABSTRACT

Background
Primary care consultation data are an important source of information on morbidity prevalence. It is not known how reliable such figures are.

Aim
To compare annual consultation prevalence estimates for musculoskeletal conditions derived from four general practice consultation databases.

Design of study
Retrospective study of general practice consultation records.

Setting
Three national general practice consultation databases: i) Fourth Morbidity Statistics from General Practice (MSGP4, 1991/92), ii) Royal College of General Practitioners Weekly Returns Service (RCGP WRS, 2001), and iii) General Practice Research Database (GPRD, 1991 and 2001); and one regional database (Consultations in Primary Care Archive, 2001).

Method
Age-sex standardised persons consulting annual prevalence rates for musculoskeletal conditions overall, rheumatoid arthritis, osteoarthritis and arthralgia were derived for patients aged 15 years and over.

Results
GPRD prevalence of any musculoskeletal condition, rheumatoid arthritis and osteoarthritis was lower than that of the other databases. This is likely to be due to GPs not needing to record every consultation made for a chronic condition. MSGP4 gave the highest prevalence for osteoarthritis but low prevalence of arthralgia which reflects encouragement for GPs to use diagnostic rather than symptom codes.

Conclusion
Considerable variation exists in consultation prevalence estimates for musculoskeletal conditions. Researchers and health service planners should be aware that estimates of disease occurrence based on consultation will be influenced by choice of database. This is likely to be true for other chronic diseases and where alternative symptom labels exist for a disease. RCGP WRS may give the most reliable prevalence figures for musculoskeletal and other chronic diseases.

Keywords
consultation and referral; databases; musculoskeletal disease; prevalence.

INTRODUCTION

Estimating the proportion of a population with a particular disease (the prevalence) provides the basis for the public health functions of needs assessment (how many people may require care for the particular disease), monitoring of changes in disease occurrence over time, and investigating differences in disease frequency between geographical areas and sociodemographic groups. As a defined number of people are registered with a general practice for the provision of health care in the UK system, prevalence estimates using self-reported data from national surveys in the UK can be used to calculate the likely numbers of people with a particular disease in the practice. However, continual recording of the reasons for consultation (a longstanding feature of British general practice) provides an alternative method of directly estimating morbidity prevalence in general practice populations and of providing prevalence estimates for symptoms and conditions which have not been studied in population surveys. The increasing computerisation of practice records has enhanced these opportunities. There are, however, a number of drawbacks to using consultation data to estimate morbidity prevalence in this way. First, not all morbidity will be presented to health services.
How this fits in

Primary care consultation data are an important source of information on the patterns and occurrence of morbidity. Large national consultation databases are popular sources of such data, and should give the most reliable prevalence figures because of the training given to practices in recording morbidity although it is not known how reliable such figures are. Using musculoskeletal conditions as an example, this paper shows the considerable variation in persons consulting prevalence figures derived from different databases. Instructions given to healthcare professionals in the recording of morbidity strongly influence the recorded prevalence and researchers and health service planners need to be aware of these differences.

the GP — consultation data will always be about ‘morbidity for which primary health care is sought’ and the extent to which it reflects the true population prevalence of a condition will vary with the condition. Second, longstanding morbidity, for which active health care is not regularly needed or sought, may appear at some point in consultation records, but not necessarily in the records for the time period chosen to estimate prevalence figures. Conditions which are episodic, or that require frequent consultation, may occur more than once in any single individual’s record during such a time period. The measure of occurrence will thus be affected by the length of the time period chosen for scrutiny and whether repeat consultations are included. Third, the consultation entry will reflect the individual GP’s choice of label or diagnosis and of what is considered important in complex consultations — in the absence of any standardisation, this will vary considerably between practitioners and across different morbidities. Fourth, the quality of recording of the reason for consultation varies between practices and also between morbidities. Despite all these concerns, the idea of using ‘persons consulting prevalence’ (the proportion of people on the practice register who consult at least once with a particular morbidity during a specified course of time such as 1-year) as a measure of true population prevalence is a popular one. As a result, national primary care consultation databases are used increasingly as sources of information on morbidity occurrence. Although such routinely collected primary care consultation data is a valuable resource for researchers and health service planners, the limitations of such data need to be understood to ensure it is used appropriately and that healthcare decisions based on it are reliable.

Consultation prevalence figures in the UK have been based on the Morbidity Statistics from General Practice national studies (the most recent being the fourth, MSGP4, in 1991–1992), the General Practice Research Database (GPRD) and the Royal College of General Practitioners Weekly Returns Service (RCGP WRS). The assumption is that these large national databases, using practices trained in the electronic recording of morbidity, will give the most reliable prevalence figures. However, few studies have compared prevalence rates derived from these different national databases.

Musculoskeletal conditions are a common reason for consultation. Fifteen per cent of the population in MSGP4 had a recorded musculoskeletal problem during the study year and one in five of all recorded consultations was a musculoskeletal problem. The umbrella term of musculoskeletal conditions includes both chronic conditions, which carry a specific diagnosis label such as osteoarthritis and rheumatoid arthritis, and also symptom-type labels such as ‘joint pain’ and ‘soft tissue disorder’ which may be episodic or acute in nature. Consultation prevalence figures for musculoskeletal conditions have previously been gathered from MSGP4.

As the first step in updating the estimates of consultation incidence and prevalence rates for musculoskeletal disease, and to examine their patterns over time, the aim of this study was to assess the consistency of consultation prevalence rates from four general practice databases. Specifically, the objective was to compare the annual persons consulting prevalence rates for musculoskeletal conditions overall, and for rheumatoid arthritis, osteoarthritis and arthralgia in these databases.

METHOD

The databases

The data sources were: i) MSGP4; ii) the RCGP WRS; iii) GPRD and iv) the Consultations in Primary Care Archive (CiPCA). Three (MSGP4, RCGP WRS, GPRD) are national UK databases of computerised general practice consultation records. The fourth (CiPCA) is a database of consultations from practices within North Staffordshire, UK.

i) MSGP4 was the fourth (and last) National Survey of Morbidity in General Practice conducted by the RCGP, the UK Office of Population and Census Surveys, and the Department of Health and coordinated by the RCGP Birmingham Research Unit. The survey took place between 1 September 1991 and 31 August 1992 and included all consultations during that period at 60 general practices in England and Wales. Practice staff were trained in morbidity data recording prior to the start of the survey. Doctors and nurses were asked to provide a diagnostic morbidity code for each problem consulted with during a contact. In particular, they were specifically encouraged to
record a diagnosis at each contact rather than symptoms (for example, to record asthma rather than wheeze).

ii) The RCGP WRS database, also coordinated by the RCGP Birmingham Research Unit, holds morbidity data from consultations routinely collected from participating practices. For this study, annual persons consulting rates are derived for the year 2001 (the likely year of the next MSGP survey if it had been conducted) based on the 38 practices in England and Wales able to submit the requisite annual data. Over half of these had contributed to the 1991–1992 MSGP4 survey.

Doctors and nurses are asked to enter a predominant symptom code only if no clear diagnostic code can be allocated with reasonable probability. The morbidity encountered at each consultation is required to be entered.

iii) The GPRD is a computerised database of general practice consultation data within the UK. Practices are required to reach a set standard in recording before contributing data and a series of recording guidelines are set down. Morbidity codes only have to be entered under three circumstances: at first diagnosis, if the event results in a change of treatment, or if another significant event, such as a referral, occurs. A chronic condition only needs to be recorded on the first occasion a new drug is prescribed. Symptom codes are allowed if seen as appropriate.

The correctness of recorded morbidity codes assigned to patients in the GPRD has been examined for several conditions by comparison with external information such as hospital information. These conditions include psychoses, venous thromboembolism, bowel disease and pressure ulcers. For this study, data for 2001 was used when 227 practices contributed data for the whole year. To allow comparison with the MSGP4, prevalence figures were also calculated for 1991 (157 practices).

iv) CiPCA is a database of consultations from nine general practices in North Staffordshire, UK. Data for 2001 was again used. The practices are part of the Keele GP Research Partnership and, as such, have regular cycles of training, assessment and feedback in the quality of their computerised morbidity coding. Current estimates show that 93% of doctor contacts at these practices are given a morbidity code.

Consultations are typically recorded in the UK using the Read Code morbidity coding system. Read Codes are a hierarchy of morbidity, symptom and process codes which become more specific further down the hierarchy. In MSGP4 and RCGP WRS, these Read Codes have been mapped to ICD9 (International Classification of Disease revision 9) codes for presentation of prevalence and incidence of disease. However, data from the MSGP4 were analysed here based on the original Read Code format.

Statistical analysis

Annual persons consulting prevalence rates were calculated per 10,000 registered persons aged ≥15 years. To be counted as a prevalent case, patients had to have one or more recorded consultations during the 12-month period with the specified Read Code. Patients who consulted more than once with that code were only counted once. Prevalence rates for any musculoskeletal condition were calculated first. The specified codes were all codes within Chapter N “Musculoskeletal and connective tissue diseases” of the Read Code system (equivalent to ICD9 codes 710–739). Then, prevalence rates for three chosen morbidities were derived. Rheumatoid arthritis was chosen to represent a specific disease category. Read codes starting N04 (‘Rheumatoid arthritis and other inflammatory polyarthropathy’), equivalent to ICD9 codes beginning 714, were identified. Osteoarthritis and arthralgia were chosen to represent the contrast between disease and symptom-type labels for a similar condition. The Read codes for osteoarthritis were those starting N05 (‘Osteoarthritis and allied disorders’), equivalent to ICD9 codes beginning 715. The Read codes for arthralgia were those starting N09 (‘Other and unspecified joint disorders’) and ICD9 codes beginning 719. Prevalence rates were calculated for adults (aged 15 years and over) from the raw data for the MSGP4, GPRD and CiPCA. Due to the method of collating data within the RCGP

| Table 1. Age and sex of the study populations (aged 15 years and over). |
|------------------|------------------|------------------|------------------|------------------|
| Number of persons | 376307*          | 780354           | 1419667          | 264076          |
| Female (%)        | 51.6             | 51.3             | 51.1             | 51.0             |
| Male (%)          | 48.4             | 48.7             | 48.9             | 49.0             |
| Age, years (%)    |                  |                  |                  |                  |
| 15–44             | 55.5             | 51.5             | 48.9             | 53.1             |
| 45–64             | 26.3             | 28.5             | 30.9             | 29.8             |
| >65               | 18.2             | 20.0             | 20.2             | 17.1             |
|                  |                  |                  |                  |                  |
| Number of person-years |              |                  |                  |                  |

*Person-years.
WRS, this was not possible for this database and prevalence rates were extracted or calculated from the annual prevalence report. Fifteen was the minimum age chosen because of the age groupings used in the RCGP WRS report.

The databases have slightly different definitions of the denominator population. The denominator population for MSGP4 was derived in terms of person years, so a person registered for half the year would add half a person year to the denominator population. In the GPRD, only patients registered throughout the 12-month period could contribute to the numerator or denominator. The denominator populations for CiPCA and RCGP WRS were the registered mid-year or end of year populations depending on practice. Persons consulting rates per 10 000 patients were produced for each database by sex and by age group. To take into account any difference in the age–sex structure of the populations covered by the databases, an overall age–sex standardised prevalence rate (with 95% confidence interval [CI]) was determined. The standard population used was that of England and Wales in 2001 (Office of National Statistics).

A further analysis calculated the percentage of all musculoskeletal consultations that were for rheumatoid arthritis, osteoarthritis and arthralgia.

RESULTS

The study populations were similar in terms of the percentage of the adult population who were female, and similar to that of England and Wales as a whole (Table 1). However, RCGP WRS and MSGP4 adult populations were younger than that of England and Wales as a whole, while the CiPCA adult population were older.

Table 2 shows the age-sex standardised persons consulting prevalence rates and age-specific rates for musculoskeletal conditions, rheumatoid arthritis, osteoarthritis and arthralgia. A consistent feature for all morbidities was that the GPRD 1991 prevalence rates were far lower than those derived from the other databases.

| Table 2. Persons consulting prevalence rates per 10 000 people aged 15 years and over of musculoskeletal conditions, rheumatoid arthritis, osteoarthritis and arthralgia. |
|---------------------------------|---------------|-------------|-------------|-------------|-------------|
| All musculoskeletal              | MSGP 91/92    | GPRD 1991   | GPRD 2001   | RCGP WRS 2001 | CiPCA 2001 |
| Age–sex standardised rate (95% CI) | (1858 to 1883) | (657 to 668) | (1292 to 1303) | (1946 to 1976) | (2038 to 2102) |
| Rate ratio*                      | 1.0           | 0.4         | 0.7         | 1.0          | 1.1         |
| Age standardised rate            |               |             |             |              |             |
| Female                           | 2122          | 751         | 1479        | 2233         | 2329        |
| Male                             | 1600          | 568         | 1101        | 1668         | 1791        |
| Rheumatoid arthritis             |               |             |             |              |             |
| Age–sex standardised rate (95% CI) | (42 to 46)  | (6 to 7)    | (21 to 22)  | (39 to 44)  | (31 to 41)  |
| Rate ratio*                      | 1.0           | 0.2         | 0.5         | 1.0          | 0.8         |
| Age standardised rate            |               |             |             |              |             |
| Female                           | 60            | 9           | 30          | 58           | 46          |
| Male                             | 27            | 4           | 13          | 24           | 25          |
| Osteoarthritis                   |               |             |             |              |             |
| Age–sex standardised rate (95% CI) | (419 to 433) | (37 to 40)  | (162 to 166) | (270 to 283) | (221 to 244) |
| Rate ratio*                      | 1.0           | 0.1         | 0.4         | 0.6          | 0.5         |
| Age standardised rate            |               |             |             |              |             |
| Female                           | 526           | 47          | 206         | 347          | 288         |
| Male                             | 318           | 29          | 120         | 200          | 172         |
| Arthralgia                       |               |             |             |              |             |
| Age–sex standardised rate (95% CI) | (268 to 279) | (119 to 124) | (387 to 393) | (425 to 440) | (388 to 420) |
| Rate ratio*                      | 1.0           | 0.4         | 1.4         | 1.6          | 1.5         |
| Age standardised rate            |               |             |             |              |             |
| Female                           | 318           | 146         | 454         | 501          | 457         |
| Male                             | 225           | 96          | 321         | 358          | 347         |

*MSGP 91/92 arbitrarily selected as reference category.
Musculoskeletal conditions

The persons consulting rates of musculoskeletal conditions followed a similar trend for all of the databases (Figure 1). Prevalence increased steadily over the age groups. CiPCA, RCGP WRS and MSGP4 showed similar prevalences for each age group. Overall annual prevalence of musculoskeletal conditions for adults aged ≥15 years was around 2000 persons per 10 000 based on these three databases; however, the GPRD 2001 prevalence rate was only around two-thirds of this figure. The prevalence rate in males was around 75% of that in females (range across databases 74–77%).

Rheumatoid arthritis

MSGP4 and RCGP WRS gave the highest adult prevalence rates for rheumatoid arthritis at around 44 per 10 000, CiPCA rates were similar, and all three gave rates about twice those of GPRD 2001 (Supplementary Figure 1). Prevalence rates for females were consistently about twice as high as for males.

Osteoarthritis

Similar overall persons consulting prevalence rates for osteoarthritis were derived from the CiPCA and RCGP WRS databases (230–280 per 10 000 people aged ≥15 years). The overall MSGP4 prevalence was over 50% higher than the prevalences from these databases. By contrast, GPRD 2001 prevalence was two-thirds that of CiPCA and RCGP WRS, and less than half that of MSGP4. The prevalence rate of osteoarthritis increased sharply after age group 25–44 years (Figure 2). Prevalence for males was around 60% that of females (range across databases 58–62%).

Arthralgia

Similar adult persons consulting prevalence rates for arthralgia were derived from each of the CiPCA, RCGP WRS and GPRD 2001 databases, 400 per 10 000 people aged 15 and over. The MSGP4 prevalence rate was only two-thirds that of the other databases. There was a sharp increase in prevalence rate in all four databases for the age group 45–64 years compared to the younger age groups (Figure 3). The prevalence in males was around 70% of that in females (range across databases 66–76%).

Percentage of all musculoskeletal consultations that are rheumatoid arthritis, osteoarthritis or arthralgia

In MSGP4 one-fifth of all Chapter N consultations in persons aged ≥15 years were recorded as osteoarthritis (Table 3), increasing with age to 40% of all musculoskeletal consultations in the ≥75 years age group; this contrasted with only one-tenth recorded as arthralgia in the same database. The percentages were reversed for GPRD 2001, which had 22% of Chapter N consultations recorded as arthralgia and 9% as osteoarthritis. The percentages for CiPCA (arthralgia 15%, osteoarthritis 10%) and
RCGP WRS (arthralgia 18%, osteoarthritis 12%) were similar to each other. The GPRD 1991 figures differed from the other databases (arthralgia 13%, osteoarthritis 4%).

The percentages of all musculoskeletal consultations recorded as rheumatoid arthritis were around 4% for CiPCA, RCGP WRS and MSGP4, twice that of GPRD 2001.

**DISCUSSION**

**Summary of main findings**

The annual persons consulting rates for musculoskeletal conditions were compared across three national general practice consultation databases and one regional database. Trends across the age and sex groups were similar for all databases. However, the actual prevalence of disease consultation varied considerably.

MSGP4 gave the highest persons consulting prevalence rates for osteoarthritis, but lower prevalence rates for the less specific diagnosis of arthralgia. This is most likely explained by differences in recording practice, with increased use of diagnostic codes in MSGP4 relative to symptom-type codes. The use of arthralgia and osteoarthritis as alternatives is most likely to occur in patients over the age of 50 years and it is in the 45–64 years age group where the divergence between databases in prevalence rates becomes apparent.

The GPRD 2001 data underestimated prevalence rates for musculoskeletal conditions in general, and for the specific disease categories of rheumatoid arthritis and osteoarthritis. This is likely to reflect the fact that, in the GPRD, not every consultation for a chronic problem has to be coded, and that there are probably a number of ‘invisible’ people who are consulting with chronic disease during the year but are not recorded as doing so. By contrast, the recording of arthralgia in GPRD 2001 was consistent with that from CiPCA and RCGP WRS. The low prevalence rate of rheumatoid arthritis in the GPRD may give most cause for concern as this is a specific disease with a definite diagnosis.

The persons consulting prevalence rates for two minor acute morbidities (chicken pox and hayfever) have been shown before to be consistent between GPRD and MSGP4. However, rates for two chronic conditions (diabetes and asthma) were lower when based on the GPRD. A further study compared the GPRD and MSGP4 and showed reasonably similar rates for most of 11 respiratory diagnoses. Both of these studies suggested differences between the two databases were due to the difference in recording instructions given to contributing GPs.

MSGP4 has been used as the gold standard to assess the face validity of prevalence rates derived from the morbidity records of general practices in order to assess quality of recording. However, this study suggests that comparisons of rates to MSGP4 may not be helpful because differences in prevalence rates may be due to differences in recording practice rather than under or over-recording of morbidity.

Persons consulting prevalence rates for musculoskeletal conditions and osteoarthritis were reported to have increased between MSGP3 (1981–1982) and MSGP4, but had fallen for rheumatoid arthritis. The GPRD has shown increased prevalence rates of disease over time for several conditions. However, these results suggest that identifying trends in prevalence of musculoskeletal disease based on changes in
consultation rates either using the two GPRD databases or comparing MSGP4 with GPRD 2001 is not a reliable reflection of true changes in consultation prevalence.

The RCGP WRS and CiPCA databases gave comparable prevalence rates. RCGP WRS is a national database, consisting of practices with training and long experience in coding morbidity, and, as such, appears to represent the most accurate current morbidity patterns based on data collected as patients consult. RCGP WRS may also allow investigation of trends over time from 1994 when the database was first computerised. CiPCA is limited by its local focus, based entirely in North Staffordshire. However, the results shown here suggest that CiPCA is a good local epidemiological resource and that if effort is put in to ensure good quality data collection, then local databases can give comparable prevalence rates for all but the rarer morbidities to larger national databases.

Limitations of the study
There are general problems in determining consultation prevalence. The derived prevalence estimates in all databases are likely to be underestimates of actual consultation prevalence as every problem discussed during a single encounter may not be coded and many people with musculoskeletal morbidity may not consult their GP with the problem.

Implications for future research and clinical practice
Primary care consultation data has many potential uses. Local databases, for example, have potential value in measuring the local burden of disease and healthcare utilisation, evaluating trends over time, and allowing comparison with national data. Primary care trusts and health planners are already using data collected from their local practices, and they are making increasing use of national databases in the same way that they use other national sources of health data (for example, the Census, health surveys) to construct estimates of need and provision. Unlike the survey and Census data, however, the meaning and validity of consultation data have not been well rehearsed in the public health and policy arenas. The instigation of the Quality and Outcomes Framework (QOF) may mean that prevalences of the included morbidities in particular are likely to be used more extensively by primary care trusts and regional planners for health service planning and that the electronic patient record will become a major resource for describing the occurrence of these specified conditions. However, they will not be directly comparable to those derived from the consultation databases examined in this paper as prevalences reported from the QOF are lifetime prevalences and not annual persons consulting prevalence rates. Although most of the included morbidities have objective diagnoses (such as diabetes) and may not be affected by the coding practice of GPs, others, such as asthma, may well be. Variations in regional prevalence may, thus, continue to reflect variations in diagnosis and labelling practice as well as sociodemographic characteristics of the populations. Furthermore, many conditions, including musculoskeletal morbidity, are not included in the QOF, and the quality and reliability of routine recording of these may be slower to improve.

The current study has found considerable variation in estimated prevalence rates for musculoskeletal conditions, and researchers and health service planners should be aware that prevalence rates of consultation for disease will be influenced by the database used to generate them, as well as the methods used to calculate the prevalence rates. All databases using routine consultation coding will reflect GP variation in the application of diagnostic criteria and in the use of labels. Databases cannot ‘enforce’ diagnostic criteria since this does not reflect clinical practice. However, they can influence coding practice, and the RCGP WRS format of encouraging diagnostic coding, but allowing symptom codes if a clear diagnosis cannot be made with reasonable probability, appears to lead to more consistency. Other influences, such as the QOF, may stimulate further uniformity in the future.

The differential effect of coding requirements within the MSGP4 and GPRD is likely to extend to conditions other than musculoskeletal conditions that are also chronic in nature or that have alternative symptom-type labels. Due to the problems of interpretation of prevalence rates derived from MSGP4 and the GPRD, the RCGP WRS may give the most reliable prevalence figures for musculoskeletal disease, and this may be true for other chronic diseases also.

Supplementary information
Additional information accompanies this article at http://www.rcgp.org.uk/bjgp-supinfo

Funding body
CiPCA is funded by the North Staffordshire Primary Care Research Consortium and the National Coordinating Centre for Research Capacity Development (NCCRC). Use of the GPRD is funded by a Medical Research Council equipment grant. The RCGP WRS is funded by the Department of Health.

Ethics committee
Ethics approval for CiPCA was gained from North Staffordshire Research Ethics Committee (LREC Project 03/04)
Competing interests
Kelvin Jordan, Mark Porcheret, Umesh T Kadam and Peter Croft all work in the Primary Care Musculoskeletal Research Centre, Keele University, which is responsible for the collection and maintenance of the CiPCA database.
Douglas Fleming is Director of the RCGP Birmingham Research Unit which coordinates the RCGP WRS database.

Acknowledgements
The authors thank Helen Ogden for the organisation of the CiPCA database, the Informatics team in the Primary Care Musculoskeletal Research Centre for collecting the data and all the general practices who contribute to CiPCA. We are also grateful for the contribution to the RCGP WRS of the collaborating sentinel practices.

REFERENCES