# Case studies

# Integrating data to facilitate clinical research: a case study

Ilana N Ackerman BPhysiotherapy (Hon)
PhD candidate, Centre for Health, Exercise and Sports Medicine, School of Physiotherapy, University of Melbourne and Centre for Rheumatic Diseases, Royal Melbourne Hospital, Parkville, Victoria, Australia

Richard H Osborne BSc PhD
Senior Lecturer, Centre for Rheumatic Diseases, Royal Melbourne Hospital, University of Melbourne, Parkville, Victoria, Australia

## ABSTRACT

The integration of routine clinical administrative activities into ongoing rigorous clinical research poses challenges for both clinicians and researchers. This case study describes the development of a responsive database system used to facilitate comprehensive longitudinal research into the outcomes of patients waiting for hip and knee replacement surgery in a large public teaching hospital. The initial research procedure was paper-based, with manual patient matching and data entry. This process was time-consuming and associated with substantial risk of error and omissions, necessitating the design of a better system. An integrated database system was designed to receive daily electronic updates of the orthopaedic waiting-list and scheduled clinic and surgery dates. Using readily available software (Microsoft Access), new patients were identified through specifying inclusion and exclusion criteria which allowed rapid and complete

recruitment at time of entry to the waiting-list. The integrated system specified the appropriate timing of multiple follow-up assessments, provided prompt information on recruitment for reporting purposes and integrated multiple linked research projects within one database. Seamless exporting of data to statistical programs for analysis was also enabled. This simple integrated approach facilitated efficient execution of a longitudinal study from recruitment to statistical analysis while maximising confidentiality and minimising resources required. This case study describes the development and design of a simple system which could be easily adapted for database management in hospital or clinic-based settings according to local requirements.

**Keywords/MeSH terms**: arthroplasty – replacement, data linkage, database management systems

## Background

Custom-designed databases are essential for clinical and research activities across a variety of healthcare settings. These databases can integrate information from a range of sources, which can then be used for both clinical and reporting purposes. Like the comprehensive health information systems described by Nenonen *et al*, such databases can link administrative, clinical and statistical information for use by a range of stakeholders.[1] Readily available health information can be exported through conversion to a variety of data formats, such as text files, and can form the basis

of the database design.[2] The use of integrated databases in primary care has been described previously in fields such as cardiac surgery and diabetes.[3,4] However, details regarding the specific design and features of such databases are not readily available in the literature. This type of practical information would be valuable for clinicians and researchers who wish to design database systems tailored to their particular requirements. This case study describes a simple integrated database system designed to facilitate clinical research. A step-by-step approach is given to highlight

the ease with which this type of database can be developed and implemented.

Our need for an integrated database system arose from our clinical research into the outcomes of patients undergoing primary joint replacement. In Australia, like most developed countries, total knee (TKR) and hip (THR) replacement surgery for osteoarthritis has been increasing each year due to the ageing population and the growing prevalence of osteoarthritis.[5,6] As demand for these procedures rises, lengthy waits for surgery are likely. In December 2002, we commenced a study to document the well-being of patients entering a hospital orthopaedic waiting-list for TKR and THR and to closely monitor health status before and after surgery. For this study to produce a generalisable assessment of people undergoing joint replacement surgery, it was important to capture a representative sample of the orthopaedic waiting-list. This required the prompt identification of new patients entering the list and the ability to track each patient's progression through pre- and post-operative periods. The potential for inefficiencies and omissions associated with the manual database system were identified and appropriate steps were taken to design a system which met the requirements of rigorous research. In this paper we report the design of a simple integrated database system which links several hospital administrative datasets with straightforward data management for patient ascertainment and monitoring.

# Identifying the need for an integrated database system

This study was conducted at the Royal Melbourne Hospital, a large public teaching hospital. Access to patient information from the Hospital's orthopaedic waiting-list was approved by the Melbourne Health Human Research Ethics Committee. At the commencement of the study, recruitment of potential participants involved manually scanning the orthopaedic waiting-list, which was only available to the researchers in a lengthy printed format. To recruit potential participants, updated copies of the complete orthopaedic waiting-list were required weekly to identify patients who had recently been added to the list. Obtaining this document required the assistance of staff from other departments, who were external to the research team. Manual screening of the list was time-consuming, as waiting-lists for each surgeon needed to be reviewed separately and patients awaiting joint replacement were not demarcated from patients waiting for other orthopaedic procedures. Once a new patient was identified, a second procedure was

required to obtain their contact details through the Hospital's electronic patient record system. This information was then manually entered into a research database. Data such as scheduled pre-admission clinic and surgery dates could also be obtained from the printed waiting-list, but determining any changes to these dates was laborious and required comparison with previous printouts. Changes in surgery dates are common due to postponement or cancellation of procedures. The substantial labour costs and risk of errors and omissions using this series of manual procedures suggested that a comprehensive longitudinal study might not be feasible, particularly as the study was intended to run for five years and would potentially involve a large number of participants. For successful conduct of this research, it was imperative that the various hospital databases be integrated and linked to the research database system.

# Integrating hospital databases

The initial step in designing an appropriate system was to determine the type of information required from the Hospital's patient record database, which contained both patient information and the orthopaedic waiting-list in separate interfaces. Information required for the study included: unique hospital unit record number, patient name and contact details, date of birth, date of entering the waiting-list, procedure details, pre-admission clinic date (if already scheduled) and surgery date (if already scheduled). Discussions with the Hospital's information technology (IT) department revealed that patient information and waiting-list information could be extracted from different locations within the Hospital's patient record database and combined into a single text file for the purpose of the study. This was achieved using Scribe software (Scribe Report Generator, PR Integration Pty Ltd). As patients on the orthopaedic waiting-list were also assigned hospital-administered procedure codes (113 for hip surgery and 114 for knee surgery), it was possible to limit the data extraction to these procedures and exclude patients waiting for other types of surgery. As the paper-based waiting-list contained the full list of all patients waiting for orthopaedic procedures, the advantage of this was twofold. The identification of new patients entering the waiting-list for THR or TKR would be streamlined, and access to the names of patients waiting for other types of procedures would not be possible, thereby maximising the privacy of these patients.

Next, the IT department programmed an auto-report to retrieve the information specified above and export it to a tab-delimited text file located on

the Hospital's server, accessible only by the researchers. The auto-report is updated daily at 2.30 a.m. This means that updated information regarding pre-admission clinic and surgery dates is always available. Once patients receive their surgery, their details are no longer available in the text file.

# Software used for the research database

Software available at both the Royal Melbourne Hospital and the University of Melbourne was used in this study. Relational database software (Microsoft Access) was used to create a patient ascertainment and study management database. This software has several features that made it useful for this setting, including:

- the ability to import external tab-delimited text files
- the ability to link separate sets of data (contained in tables) within a single database
- the ability to 'update' existing data tables from information contained in external data files
- a sophisticated 'query' feature which can select or exclude individuals based on a range of researcher-specified characteristics
- efficient data entry features using purpose-designed forms
- commonly used statistical software packages, such as Statistical Package for the Social Sciences (SPSS Inc.), are able to read Access tables and queries for instantaneous data extraction, filtering and concatenation
- the availability of training and technical support.

# Database design

## Adding new patients to the database

The next stage involved the addition of new patients entering the waiting-list to the research database. This required the design of a sequence of queries, as illustrated in Figure 1. The first of these was a 'select query' to identify all patients on the electronic waiting-list for either THR or TKR. This query was also used to exclude surgical procedures which did not meet our inclusion criteria for the study (such as bilateral joint arthroplasty or hemiarthroplasty). This meant that patients waiting for these procedures would not be identified.

A second select query (acting on the results of the previous query) then identified all patients who were not already registered in the database (matched using unique hospital unit record numbers). This ensured that patients could only enter the study once (patients waiting for subsequent surgical procedures were excluded according to the study protocol).

An 'append' query was then used to add the details of new patients on the waiting-list for TKR or THR (who were not already in the study) to the research database. This information included unique hospital unit record number, name, contact information, date of birth and procedure details. These patients were subsequently allocated a unique study identification number and could be recruited into the study.

# Database-supported management of the longitudinal study

The design of the longitudinal study also required the collection of data from study participants at pre-determined points after entry to the orthopaedic waiting-list. These included pre-admission (the point at which patients were scheduled to visit the pre-admission clinic for pre-operative investigations, approximately two to four weeks before surgery) and fixed intervals after surgery (3 weeks, 3 months and 12 months). At these times, questionnaires were mailed to participants, and for selected participants, movement laboratory appointments were also scheduled.

The database system was designed to alert the researchers to the actions required for each time point. This was achieved by designing a 'select query' (acting on the waiting-list text file) which was used to identify the allocation of pre-admission clinic and surgery dates for study participants. This information was then appended to the database table and a pre-admission questionnaire could be sent. Once a participant's surgery date had been scheduled, an 'update' query was used to calculate the 3-week, 3-month and 12-month post-operative dates and automatically enter these into the database. Separate select queries were then used to identify, at any given time, which participants required 3-week, 3-month and 12-month questionnaires to be sent. This process is illustrated in Figure 2.

If a participant's surgery date was postponed, this was identified using the 3-week query, which would state the revised surgery date (from the waiting-list text file) so that the database table could be amended appropriately. This ensured that post-operative questionnaires were not sent to participants who had not yet received their surgery. Further select queries for each time point were designed to identify participants who had not returned completed questionnaires, so that direct telephone follow-up could be undertaken. This step was crucial in minimising missing data.
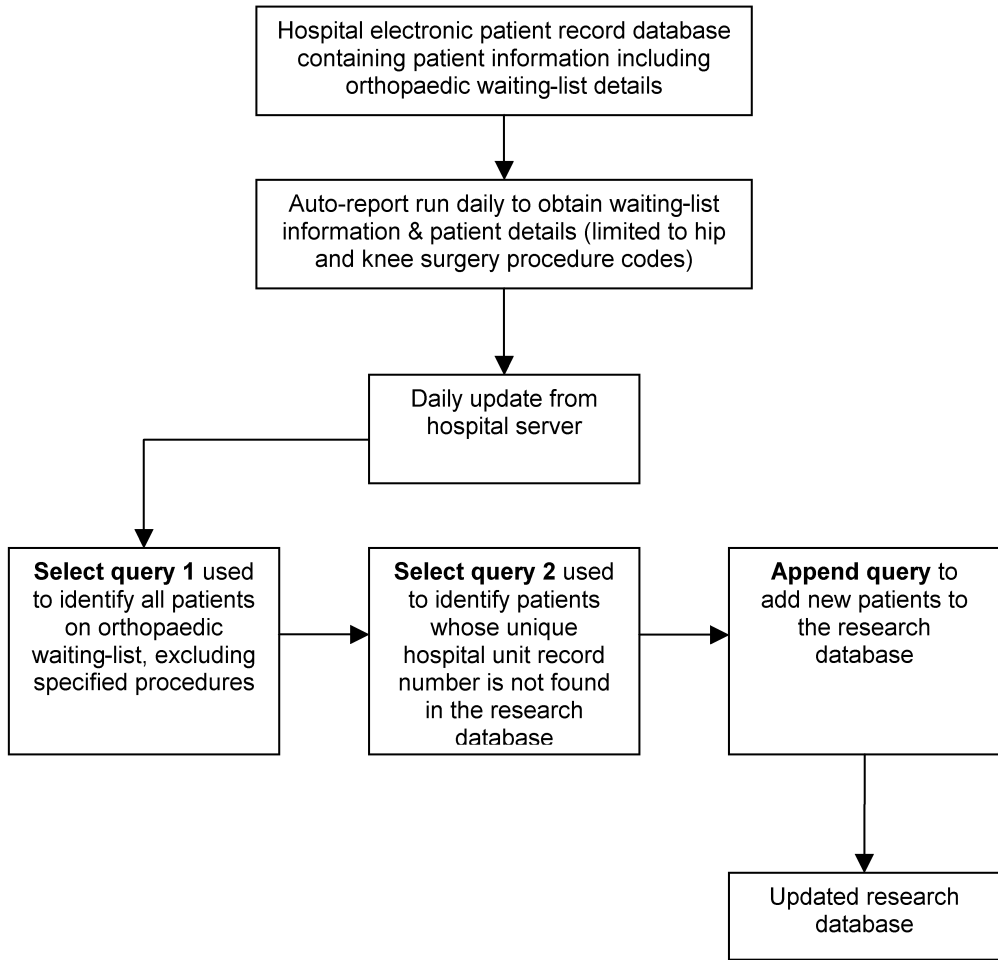
**Figure 1** Identification of new patients entering the waiting-list and addition to the research database

## Using the database for associated studies

The longitudinal study was also designed to provide a recruitment base for two associated randomised controlled trials (RCTs). One table for each RCT was established within the existing Access research database, containing identical fields for name, contact details, date of birth and procedure information, as described earlier. The three tables were linked by study identification number (subjects in each RCT retained their original study identification number). Additional fields specific to each RCT were also included in the new tables. When participants from the longitudinal study were recruited into one of the RCTs, information about these participants from the original database table was appended to the relevant RCT table using a series of queries. This eliminated the need for manual data entry. A similar set of queries, as described for the longitudinal study, was then used to calculate reassessment dates before and after surgery and identify appropriate dates to send follow-up questionnaires and contact patients about movement laboratory appointments. This example highlights the ease with which linked research projects can be managed using an integrated database system.

## Exporting data for analysis

The database system was designed to link seamlessly with SPSS Version 11.0 for data analysis. Using Access, a series of select queries was created to extract data from both the research database (containing patient information) and a data entry database (containing all questionnaire and test results) for export to SPSS. This process is illustrated in Figure 3. As both databases were linked using unique study identification numbers, data from all time points in the longitudinal study could be combined in SPSS for statistical analysis. SPSS has the ability to read Access data files and calculate instrument summary scores using a single syntax file.
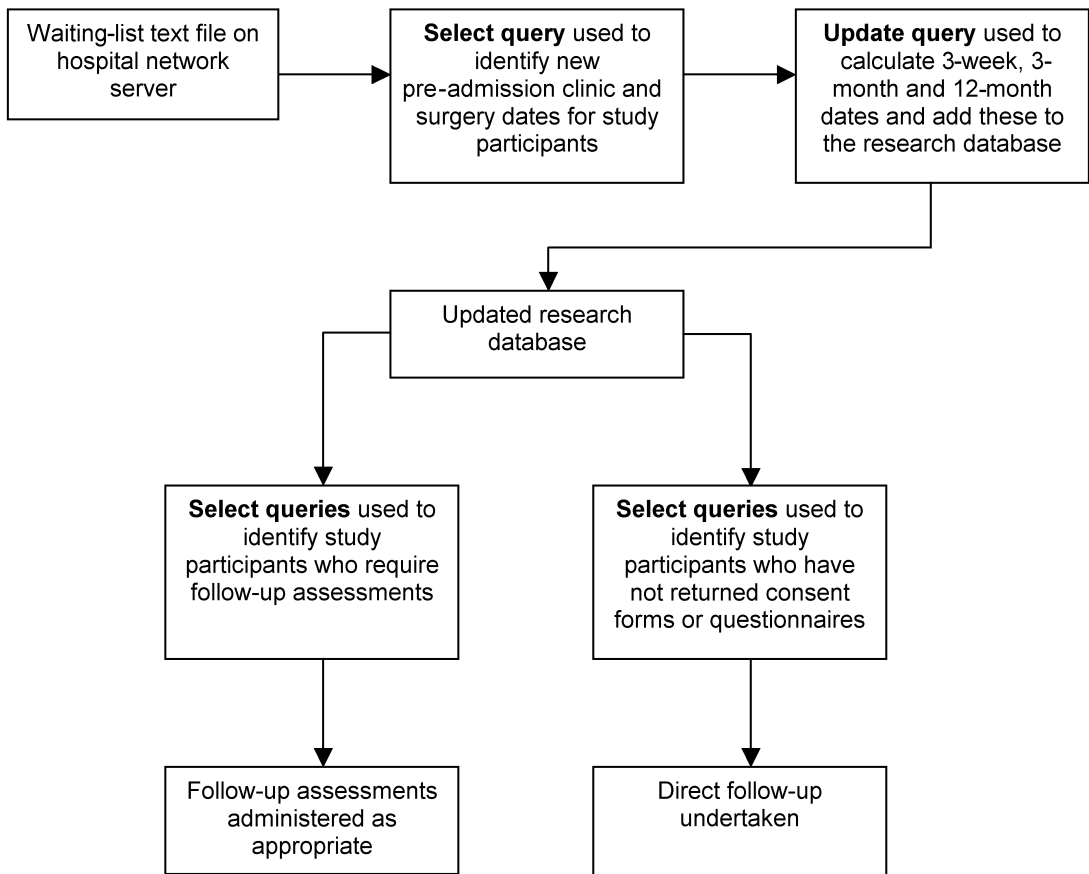
```
┌─────────────────────┐     ┌─────────────────────┐     ┌─────────────────────┐
│ Waiting-list text   │     │ Select query used   │     │ Update query used   │
│ file on hospital    │────▶│ to identify new     │────▶│ to calculate 3-week,│
│ network server      │     │ pre-admission clinic│     │ 3-month and 12-month│
│                     │     │ and surgery dates   │     │ dates and add these │
│                     │     │ for study           │     │ to the research     │
│                     │     │ participants        │     │ database            │
└─────────────────────┘     └─────────────────────┘     └─────────────────────┘
```

```
              ┌─────────────────────┐
              │ Updated research    │
              │ database            │
              └─────────────────────┘
```

```
┌─────────────────────┐     ┌─────────────────────┐
│ Select queries used │     │ Select queries used │
│ to identify study   │     │ to identify study   │
│ participants who    │     │ participants who    │
│ require follow-up   │     │ have not returned   │
│ assessments         │     │ consent forms or    │
│                     │     │ questionnaires      │
└─────────────────────┘     └─────────────────────┘
```

```
┌─────────────────────┐     ┌─────────────────────┐
│ Follow-up           │     │ Direct follow-up    │
│ assessments         │     │ undertaken          │
│ administered as     │     │                     │
│ appropriate         │     │                     │
└─────────────────────┘     └─────────────────────┘
```

**Figure 2** Use of database system to facilitate participant progression through the study

# Data protection and confidentiality

Both the research database and the waiting-list text file were password protected at server, computer and software program levels and could only be accessed by authorised research staff. The research database was stored on the Hospital's server to enable daily back-up. Data entry for questionnaire responses and movement laboratory results was performed by a data entry operator into a separate Access database using de-identified data. The data entry operator did not have access to the Access database containing patient information to further preserve confidentiality.

# Ensuring data accuracy and quality

The accuracy and quality of the research database was verified in a number of ways:

- The waiting-list text file was manually scanned to check that it contained the details of patients who had entered the list following the most recent orthopaedic outpatient clinic.
- New patients added to the database were checked against the text file to ensure that potentially eligible patients had not been missed.
- All participants without surgery dates were checked against the Hospital patient record database periodically to ensure that they were still on the orthopaedic waiting-list and had not received their surgery.
- The Hospital's electronic waiting-list was checked (using an Access query) before mailing three-week post-operative questionnaire packages so that these were not inadvertently sent to participants whose surgery had been postponed.
- Records were kept of participants whose surgery was postponed; this allowed for monitoring of the Hospital's electronic waiting-list so that updated surgery dates could be identified and post-operative questionnaire packages could be administered appropriately.
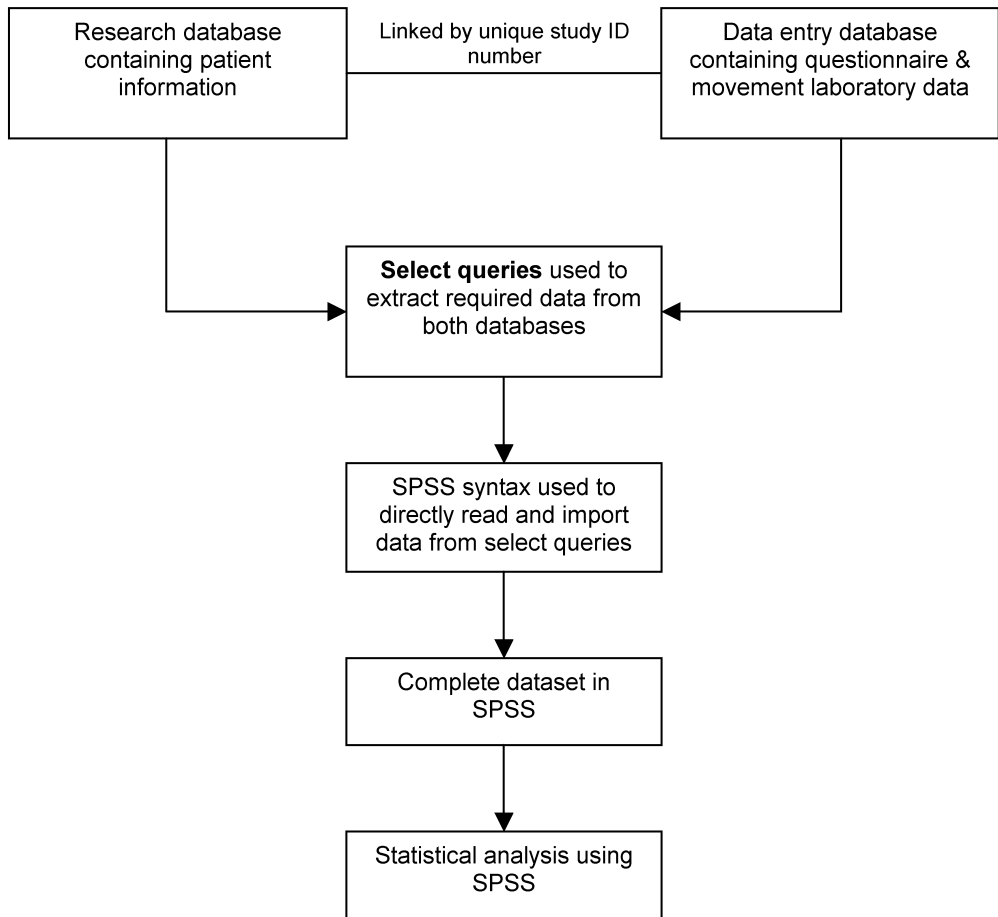
**Figure 3** Data extraction process

# Advantages of an integrated database

This simple system streamlined the management of our longitudinal study involving patients waiting for THR or TKR. Compared with traditional paper-based methods, this system permits the timely identification and recruitment of potential research participants, and the straightforward monitoring of participant progression through the study protocol. The system is operated in an uncomplicated manner by running a series of Access queries in a specified sequence. This means that the time from data entry of questionnaire responses to statistical analysis is approximately 20 minutes. Minimal manual data entry is required, reducing the introduction of erroneous information. In terms of human resources, the system requires only one person to identify new patients who might be eligible for the research study, recruit potential participants and appropriately administer follow-up assessments. The need for additional assistance from other departments, as had previously been necessary,

is eliminated. In addition to improving efficiency, the use of a sole staff member also maximises confidentiality. Efficiency is also enhanced by integrating multiple linked research projects within the one database.

The design of the integrated system ensures the smooth transfer of data from the Hospital's patient record database to the research database, and finally to statistical software for analysis. In particular, the ability to export data to statistical software easily means that preliminary data can be reviewed at any time during the five-year study. This has been useful for informing associate researchers and also allows the data to be used in the preparation of journal and conference papers during the course of the study.

A major strength of the system is the use of multiple queries, which allowed the researchers to obtain valuable information rapidly without needing to review manually large volumes of paper-based and electronic data. Queries which provided information about recruitment, participant numbers and numbers lost to follow-up were important for tracking the progress of the study, for departmental reporting and for annual reporting to the Human Research Ethics Committee.

A query which identified the number of patients who had entered the study in a 12-month period was fundamental in estimating future recruitment and projected sample size. The use of queries to identify unreturned questionnaires helped to improve retention of participants in this longitudinal study as these participants could be identified and contacted, instead of being lost to follow-up. Other queries were also useful for administrative purposes, and provided valuable information about the cohort, such as time spent on the waiting-list.

## Potential clinical applications

Although this system was designed to address our perceived research need, this integrated approach could also be adapted for clinical requirements. For example, this system could be modified effectively for use in a variety of out-patient clinics, where patients are reviewed at regular intervals after surgery or after commencing specific therapies. The database could be used to identify patients who have had recent surgery or commenced treatment and to generate clinic appointments for specified time points. It could also alert clinicians to patients who do not attend scheduled visits so that appropriate follow-up can occur. Routine post-operative investigations such as radiographs and blood tests could also be tracked and incorporated using the database. As described in the case study, the system could be used to retrieve hospital-related data from hospital databases (such as patient contact and surgery details) and link this with patient results (from investigations, medical assessments or questionnaires) in a single database. Reports could then be generated by Microsoft Access, and the data used for clinical decision making or to evaluate the performance of procedures, medications and even longer-term outcomes, such as survival of surgical prostheses. Similar systems could be designed to produce registers of patients with specific conditions and to monitor health and medical management. An Access database could also be linked to web-based data entry, so that patients could fill in questionnaires online or could provide specific health information to their general practitioner or other medical professional prior to arrival at the clinic. As health information systems vary widely across healthcare institutions, the database system presented in this case study would require location-specific modification. The design of such databases for use in primary care should also be sensitive to the needs of clinicians, and an understanding of the proposed functions of the database is essential in order to design the most appropriate system.[7]

## Limitations of the design

The database system described in the case study has several limitations related to our research setting. At the time of entry to the waiting-list, only hospital-assigned procedure codes were available to the researchers for use in our system. For comparison with clinical or research data from other settings, standardised coding of diagnoses such as International Classification of Disease (ICD) codes would be required. The lack of consistent codes has been previously identified as a problem in combining electronic health information from general practice patients.[8]

In this case study, local (hospital) unique identifiers were used to recognise new patients entering the waiting-list who were not already in the research database. We did not utilise unique national (Medicare) identifiers and did not have linkage to other hospitals, although these features would have permitted tracking of patient migration. Although this was not problematic for our research, this simplistic approach might not be feasible in clinical settings where patients might change their name over time. In these circumstances, additional methods of verification would be required, such as date of birth or alternative unique identifiers, where available.

## Conclusion

This case study shows that there is great value in exploring administrative and research data linkage options, particularly through increasing ascertainment accuracy, timely application of follow-up assessment and reduction of costs through decreased administration expenses. We have presented this case study as a step-by-step guide to illustrate the evolution of a highly valued research tool using locally available systems which could be adapted to support a variety of routine clinical activities.

## REFERENCES

1 Nenonen M, Nylander O, Rintanen H and Suominen T. What might a health information system look like? *Informatics in Primary Care* 2002;10(3):135–8.

2 Sullivan FM, Pell JP, Sweetland M and Morris AD. How could primary care meet the informatics needs of UK Biobank? A Scottish proposal. *Informatics in Primary Care* 2003;11(3):129–35.

3 Johnson T and Ventura R. Applied informatics for quality assessment and improvement. *Journal of Nursing Care Quality* 2004;19(2):100–4.

4 Morris AD, Boyle DIR, MacAlpine R *et al*. The Diabetes Audit and Research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. *British Medical Journal* 1997;315:524–8.

5 Australian Orthopaedic Association National Joint Replacement Registry. *Annual report*. Adelaide 2003.

6 March LM and Bagga H. Epidemiology of osteoarthritis in Australia. *Medical Journal of Australia* 2004;180 (Suppl.):S6–S10.

7 Davis E. Avoiding the 'dataset mentality'. *Informatics in Primary Care* 2003;11(2):49–52.

8 Morris L, Taylor MW, Hunter C and Campbell LM. Information management in primary care: delivering a strategy to improve patient care in Scotland. *Informatics in Primary Care* 2002;10(2):85–8.

## CONFLICTS OF INTEREST

None.

## ADDRESS FOR CORRESPONDENCE

Ms Ilana Ackerman
Centre for Health, Exercise and Sports Medicine
School of Physiotherapy
The University of Melbourne
Parkville 3010, Victoria
Australia
Tel: +61 3 8344 4171
Fax: +61 3 8344 3771
Email: i.ackerman@pgrad.unimelb.edu.au