# 3: How to use databases in sports medicine research

JOHN ORCHARD, GREG BLOOD

## Introduction

A database is an organised collection of related information, stored in a format that enables efficient retrieval. Anyone who is undertaking or analysing research in sports medicine will inevitably use databases, with or without realising it. The power of internet is only possible due to search engines, which are powered by massive databases containing information (in the form of key words) about millions of websites. The ability to thoroughly review a topic in sports medicine requires the use of a literature database such as Medline (PubMed) or SportDiscus. Primary research in sports medicine has been possible in the past without recording results in a database, but as the average numbers of subjects in studies increase from dozens to hundreds and thousands, databases will be essential to efficiently manage the greater amounts of information. When papers are ready for submission to journals, Citation databases such as EndNote, Reference Manager and ProCite can ease the arduous process of correctly formatting the bibliography for the various journals.

Strictly speaking, the term "database" refers to the organised collection of data (the information itself in its ordered form). However, the term database is often also used to describe the structure for the information, or even the program used to create this structure. For example, a day surgery unit may have a Microsoft Access file containing data about operations performed at the unit. The term "database" is often used to describe the program (Microsoft Access), the structure of the file written to record all of the operative details (without the data itself), or the file containing all of the data (which is the true meaning of database).

## Searching for sports medicine information on the Internet using databases

The internet now provides sports medicine researchers with instant access to a range of medical and other information. The medical area

has been one of the growth areas of the internet and there are an estimated 15 000 to 20 000 medicine related websites. In the medical field governments and associations have led the way in making available medical information. For instance, the National Library of Medicine, funded by the United States Federal Government, several years ago made freely available on the internet the Medline database.[1] The internet version of Medline was named PubMed and now provides extremely up-to-date access to information through any terminal connected to the internet *(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed)*.

The earlier periods of the internet were characterised by questionable, inaccurate and incomplete information. Whilst in some cases this is still the case, the internet can now be used as a major means of locating up-to-date and reliable information in the sports medical arena. There are an increasing number of reputable organizations and individuals publishing information on the internet. The type of reliable and current information that may now be located includes:

- bibliographies and author reference listings – many authors and departments are now publishing lists of their research or articles with reference lists (for example *http://www.physioth.unimelb.edu.au/csmre/pubs.html*)
- conference papers and abstracts – there has been a decline in printed conference proceedings and abstracts over the last ten years. Organisations such as Sports Medicine Australia now publish conference abstracts on the internet *http://www.ausport.gov.au/fulltext/2000/preoly/*
- contents pages of journals – as can be seen from Table 3.1 most sports medicine journals now publish the contents pages of recent issues on the internet. Abstracts are included for most journals.
- commercial services such as Proquest include summaries of articles from large collections and provide full text access to journals. For instance, the *American Journal of Sports Medicine* is only available in full text through this service.
- library catalogues such as the National Library of Medicine, British Library, university libraries can now be searched individually or as a conglomerate through z39·50 library gateways
- medical review websites are increasingly becoming available through the internet. One of the most heavily used evidence based medicine databases available through the internet is the Cochrane Library (*http://www.cochrane.org/*) that identifies and collates data from randomised trials and produces systematic reviews.
- several organizations have developed sports medicine gateways. An excellent gateway in the sports injury area in the Sports Injuries

**Table 3.1 Lists major sports medicine journals, their inclusion in major databases and the availability of contents pages and full text articles on the internet. Y (Yes) N (No)**

| Title | PubMed | Sport Discus | Contents | Full Text |
|---|---|---|---|---|
| The American Journal of Sports Medicine | Y | Y | Y | Y |
| British Journal of Sports Medicine | Y | Y | Y | Y |
| Clinical Journal of Sports Medicine | Y | Y | Y | Y |
| Clinics in Sports Medicine | | Y | Y | N |
| International Journal of Sports Medicine | Y | Y | Y | N |
| Journal of Athletic Training | N | Y | Y | N |
| The Journal of Orthopaedic and Sports Physical Therapy | Y | Y | Y | N |
| Journal of Science and Medicine in Sport | Y | Y | N | N |
| Journal of Sport Rehabilitation | N | Y | Y | N |
| Journal of Sports Chiropractic and Rehabilitation | N | Y | N | N |
| The Journal of Sports Medicine and Physical Fitness | Y | Y | N | N |
| Journal of Sports Sciences | Y | Y | Y | Y |
| Journal of Sports Traumatology and Related Research | N | Y | Y | N |
| Knee Surgery, Sports Traumatology, Arthroscopy | Y | Y | Y | N |
| Medicine and Science in Sports and Exercise | Y | Y | Y | Y |
| New Zealand Journal of Sports Medicine | N | Y | N | N |
| Operative Techniques in Sports Medicine | N | Y | Y | N |
| Physician and Sportsmedicine | N | Y | Y | Y |
| Physiotherapy in Sport | N | Y | N | N |
| Scandinavian Journal of Medicine and Science in Sports | Y | Y | Y | Y |
| Sports Medicine | Y | Y | Y | N |
| Sports Medicine and Arthroscopy Review | N | Y | Y | N |
| Sports Medicine Training and Rehabilitation | N | Y | N | N |

NB This Table correct as of July 2001. It should be noted that many journals are now moving to full text access.

gateway which is part of the Medline Plus Health Information service is *(http://www.nlm.nih.gov/medlineplus/sportsinjuries.html)*. This gateway only lists documents from major medical organizations.

- full text articles, papers and reports are increasingly being published on the internet by organizations and researchers
- directories of sports physicians are now being made available by sports medicine associations
- statistical information is still limited but organizations such as the National Centre for Catastrophic and Sport Injury Research are using the internet to publish statistical research
- many websites publish information on injuries suffered by professional athletes, often for the purposes of betting or participation in fantasy leagues. A review of National Basketball Association injury information on the web concluded that the majority of published information is probably accurate.[2]

The major advantage of searching the internet for sports medicine information is the huge number of websites containing information, meaning that rare but important information can be uncovered. The other advantage is that the internet may contain the most up-to-date information from a particular source. It normally takes 18–24 months for an article to appear in a refereed journal from the time it is written, whereas some authors may publish on their own websites immediately on completion of a study. The major disadvantage of searching the internet is that there is no guarantee that information will be accurate, or that searches will be complete. If information is found via a link from a reputable source, it is more likely to be accurate.

The most common way of searching the internet is by using a search engine for example Google (*http://www.google.com*), Alta Vista (*http://www.altavista.com*) Yahoo (*http://www.yahoo.com/*). Search engines are powered by databases with information on the content of websites.

In using a particular search engine it is important to understand how a search engine retrieves information and what searching features are available. A common mistake many searchers make is that they do not use suitable search terms or use phrase searching. A recent development has been the availability of subject search engines such as Scirus (*http://www.scirus.com/*) that only search the web for scientific information.

Other internet tools that can be used for locating information include listservs and forums. These tools allow the researcher to send out requests for information to other researchers with a similar interest. Unpublished data may obtained this way. There are also subject gateway websites that organise relevant websites on the internet i.e. Bandolier Evidence Based Health Care website (*http://www.jr2.ox.ac.uk/bandolier/*).

**Table 3.2 Comparison of PubMed and SportDiscus databases.**

| | PubMed | SportDiscus |
|---|---|---|
| Cost | Freely available on internet | Fee based |
| Access | Pub Med on Internet Medline on internal networks | SportDetective on the internet, SportDiscus on internal networks |
| Core Journals | Excellent coverage | Excellent coverage |
| Non core journals | Limited coverage | Excellent coverage |
| Conference Proceedings | No coverage | Good coverage |
| Books, Reports, Theses | No coverage | Good coverage |
| Currency in Indexing | Very current | Several months late |
| Level of Searching | Sophisticated | Medium |
| Full Text Links to Articles | Yes – Limited | Yes – limited |
| Marking Records | Yes | Yes |
| Downloading to Citation Databases | Yes | Yes |
| Document Supply Service | Yes | Yes |

## Searching for scientific papers using literature databases

Whilst the internet is increasingly providing access to timely and accurate medical information, literature databases such as PubMed and SportDiscus still remain the best starting point in the research gathering process. Most medical literature databases can now be accessed through the internet and have links, where available, to full text documents on the internet. The major sports medicine related databases PubMed and SportDiscus in the last few years have devoted considerable resources to ensure that they are up to date and provide instant access to information. The differences between PubMed and SportDiscus are summarised in Tables 3.1 and 3.2, which compare the databases in terms of journals covered, currency and functionality. It is strongly recommended that both databases be searched if a comprehensive search of the literature is required.

PubMed (also known as Medline) provides access to over 11 million journal citations in 4500 journals (*http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed*). All core sports medicine journals are covered as well as allied journals in fields of orthopaedics, physical therapy and biomechanics. Features that should be utilised to ensure effective use is made of PubMed includes the following examples.

- MESH Browser – this allows the searcher to select correct search terms and their subheadings. The use of this function is extremely

useful in restricting searches to the most relevant citations particularly where there has been a large amount of research published. MESH browser allows the searcher to connect clinical filters i.e. classification, epidemiology, etc. to the suitable MESH term i.e. Fractures, Stress/Epidemiology

- Limit function allows the search to be limited by specific age group, gender, human or animal studies, language and specific publication types including clinical trials, meta-analysis, randomised control trial and reviews. This function is extremely useful when searching for evidence-based medicine research.
- Clinical queries function allows searches to be restricted to four study categories – therapy, diagnosis, etiology, prognosis.
- Journal Browser allows for the listing of the latest articles in a journal.
- Cubby feature allows you to save frequently used search strategies.
- Marking, saving and downloading citations.

To limit searches to sports medicine information in PubMed commence with the search strategy – sports OR sports medicine OR athletic injuries – select these terms from the MESH Browser. If the MESH Browser does not list a suitable term then search the database using the term you know i.e. shin splints, osteitis pubis. This is frequently the case with sports medicine terminology.

Another excellent feature of PubMed is the linking of full text journal articles to citations where available. This feature is dependent on the searcher having access to the journal through personal or organization subscription.

SportDiscus managed by the Canadian Sport Information Resource Centre (SIRC) and endorsed by UNESCO as the international database for sport should also be searched. Sports medicine researchers generally do not search this database but Tables 3.1 and 3.2 highlight the fact that SportDiscus provides access to research in conference proceedings, theses, chapters in books and non-core sports medicine journals that are not covered by PubMed. Currently there are over 500 000 citations listed on SportDiscus. Access to SportDiscus is through a library network or through SIRC's SportDetective internet service. SportDiscus lists high-level research and practical information. The "advanced" level function should be used to restrict citations to original research.

The SMART database from the National Sports Medicine Institute located in London, England has over 32 000 citations from 1986 on sports medicine. Access is through the internet but there is a monthly or annual subscription fee. The relatively new internet based Sponet database produced by the Institute of Applied Training Science at Leipzig in Germany is providing access to training and sport science websites and internet documents.

Non-English sport databases that cover sports medicine that could be searched include

- Heracles, a French language database by the Institut National du Sport et de l'Education Physique in Paris
- Spolit, a German language database by the Federal Institute of Sport Science in Cologne
- Atlantes, a Spanish language database by the Latin-American Association for Sports Information.

Other medical and scientific databases that may be searched include Cumulative Index to Nursing & Allied Health (CINAHL), Excerpta Medica and Science Citation Index.

Whilst many journals now provide access to their contents through databases or the internet, there is still often the need to obtain the printed copy of the article. Organisations that can assist in this process include:

- National Library of Medicine (United States) – *http://www.nlm.nih.gov/*
- British Library (United Kingdom) – *http://www.bl.uk/*
- Canada Institute for Scientific and Technical Information (Canada) – *http://www.nrc.ca/cisti/*
- Sport Information Resource Centre (Canada) – *http://www.sirc.ca/*
- National Sport Information Centre (Australia) – *http://www.ausport.gov.au/nsic/*
- National Sports Medicine Institute (United Kingdom) – *http://www.nsmi.org.uk/*

With the improvement of access to databases through the internet and local networks, there has been a move away from librarians to medical researchers in searching databases. Whilst this situation is beneficial to researchers on the one hand, Haynes *et al* found that this situation is resulting in inexperienced searchers missing relevant citations because of inefficient searches.[1] Medical librarians should still be utilised particularly in preparing searching strategies, as well as teaching clinicians quality filtering and appraisal of the literature.[3]

## Storing references and formatting a bibliography using Citation Databases

The integration of many database products has greatly improved the management of citations for research purposes. PubMed and SportDiscus both allow the downloading of records to citation databases such as Reference Manager, EndNote and ProCite. These

programs have filters and connection files to ensure that information is correctly downloaded from the external database into the user's library. This allows the researcher to select and export relevant citations to their citation database for future use in bibliographies and reference lists. Citation databases allow you to reformat citations to meet the citation styles of hundreds of scientific journals. Formatting styles for the major journals are already included in the citation database program. Less well known journals can have their style inputted by the user. For example, EndNote version 4 arrives with the style for *Medicine and Science in Sports and Exercise* already contained within the program, but the formatting style for *Sports Medicine* must be set-up by the user.

Many authors use citation databases to manage their own published research. If there is the possibility that a scientific paper may be submitted to more than one journal, or have references added after the review process, then the use of a citation database can save hours of time and decrease the chances of an error in formatting or a mismatched reference. It is surprising that most journals in the sports medicine field, to date, do not require papers to be submitted with a citation database file. Although most journals now encourage electronic submission (such as in Word or Word Perfect format), the journal editors generally expect the authors to manage their own reference list and then proceed to edit the references within the word processing program. Submission in the future will require authors to submit both a word processor file for the text, with citations linked to a citation database, which is also supplied. The editing process will involve the editors matching the authors' references in their citation database to the journal's citation database (which presumably will be less likely to contain errors). Authors may be required to provide reference IDs such as PMID (PubMed ID). These innovations will be introduced as citation database programs become able to undertake the reference matching process automatically. They will be necessary as journals become full-text on the web, and reference formats for journals include URLs (Uniform Resource Locators, or web addresses) as compulsory fields.

There are examples of medical libraries downloading relevant citations on a select topic and creating their own internal evidence-based medicine database that can be accessed by local clients on their network.[4]

## Designing your own database within a spreadsheet or database program

Although not every sports medicine researcher will need to become a database programmer, the ability to program using a user-friendly

database such as Microsoft Access is a very useful skill for a sports medicine researcher. Many sports medicine professionals and researchers are now comfortable using word processor and spreadsheet programs – skills that were rare 20 years ago. The ability to design a basic database is a skill that may be considered rare today, but will become a standard skill in the future, as more professionals appreciate the power of databases.

The biggest advantage of a self-designed database is that it includes exactly what you want it to include for the task at hand (or the study that you are conducting). A database structure can be planned by someone who cannot program a database, and then given to a professional programmer to create. If you not only design the database yourself, but also create it, you have the added advantage of being able to modify it whenever you wish to add or extract extra information. One of the most important factors to consider whenever you elect to use a database that has been designed by someone else is the ease of exporting data. To protect their intellectual property, professional programmers or companies selling databases will lock the programming code so that it cannot be seen by the user. This means that once bought, the structure of the database cannot be changed without going back to the original programmer. However, some databases on the market do not even include the facility to export the raw data, which the user enters, to another format. This means that after entering the data, the user can only use that specific program to analyse their data. If another type of analysis is desired, using a different database program, it may not be possible if there is no export function in the original program.

The simplest form of a database is known as a "flat file", in which all of the information is stored in a single table. Spreadsheet and even word processor programs can be used to store data in a flat format. For example, operation reports could be stored in a spreadsheet table in the format of Table 3.3.

Creating a worksheet in a program such as Excel is a simple process – the user only needs to start typing and a table will be created. The program will automatically detect special formats like numbers, percentages, dates and currency amounts. Automatic or manual formatting can be used to keep the column and row widths suitable for the amount of data in each cell. Certain functions are available within spreadsheet programs to analyse data. For example, data can be sorted by operation type (alphabetically) or date. If the operative fee was included as a field, then the fees for all of the operations on a particular day or week can be totalled.

More complicated information analysis is difficult with standard spreadsheets. Continuing the current example, it would be difficult in Excel to retrieve all records of knee reconstructions using the

**Table 3.3 Example of a 'flat file' format for a database.**

| Patient surname | Patient first name | Diagnosis | Side | Date of injury | Date of surgery | Type of surgery | Hospital |
|---|---|---|---|---|---|---|---|
| Smith | Mary | ACL tear | R | 1/3/00 | 15/4/00 | ACL reconstruction using patellar tendon | St. Elsewhere's |
| Bloggs | Joe | Lateral meniscal tear | L | 11/3/00 | 15/4/00 | Arthroscopic partial menisectomy | St. Elsewhere's |
| Jones | Fred | Knee osteoarthritis | L | 1/1/96 | 15/4/00 | Knee replacement | St. Elsewhere's |

**Table 3.4 Major software programs that can be used to store databases.**

| Type of file | Examples |
|---|---|
| Spreadsheet | Microsoft Excel<br>Lotus Notes |
| Relational database program | Microsoft Access<br>Dbase<br>Fox Pro<br>Lotus Approach |
| Citation database program | EndNote<br>Reference Manager<br>ProCite |
| Injury monitoring software | Injury Tracker<br>Sport Care<br>SIMS<br>Sports Injury Manager |

patella tendon, between 1997 and 1999, where the time to surgery was less than two months after injury. Lotus Notes is a more sophisticated spreadsheet program with superior ability to perform sort and filter functions, yet with a similar ease of data entry for the unsophisticated user.

However, the most powerful form of data storage is in a type of database known as a relational database. Examples of relational databases include Microsoft Access and FoxPro, dBase and Lotus Approach. (see Table 3.4) A relational database uses many tables that are linked together by common fields. In the operations example, there would be separate tables for "Patients", "Procedure types", "Hospitals", "Surgeons" and "Injury codes" linked by key fields (Figure 3.1). The structure of a relational database is harder to picture, as an extra dimension is added, but it makes the database far more powerful. The tables in a relational database will appear to contain less

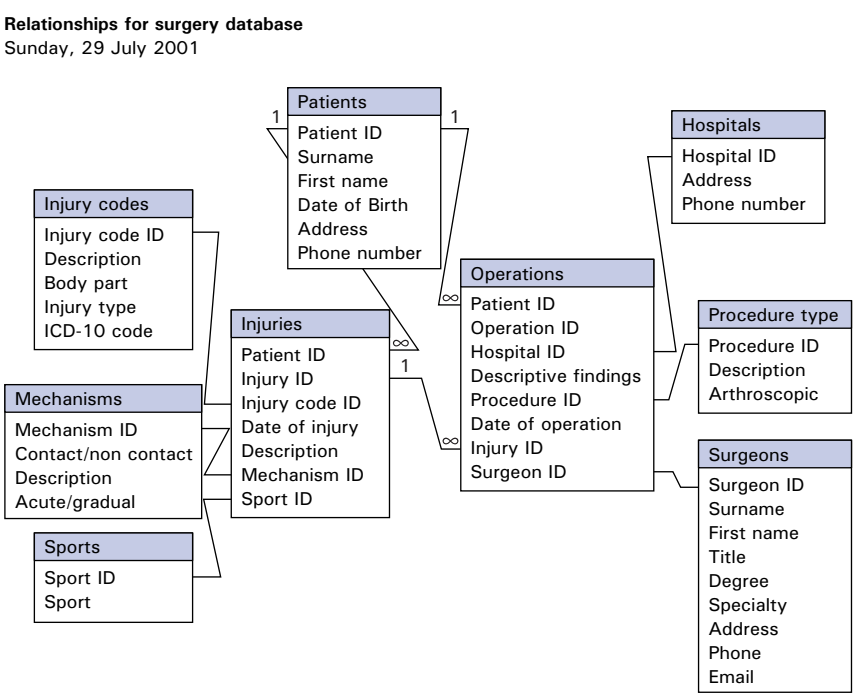


**Figure 3.1** Sample structure for a relational database to record details of surgical operations

immediately apparent information when viewed individually. The power of a relational database is realised when tables are combined to give queries.

The major advantages of a relational database are as follows.

- *Less errors*. If a commonly used entry is inputted multiple times into a flat file (e.g. Name of hospital – St. Elsewhere's) then occasionally it will be misspelt. This may cause a case to be missed during a filter operation, as it does not match the correct spelling. In a relational database, a hospital name will be entered once only, and all references to that hospital will be linked to the name record.
- *Easier updating*. When information changes (such as the address of a patient), it is changed in only one field in a relational database, which automatically updates all queries in which the field appears. In a flat file, information is often duplicated many times.
- *Smaller file sizes*. Because of the removal of duplicate information, storage of a relational database is more efficient. Forms and reports take less time to be created because of the efficient storage.

- *Power*. If the data is part of the database, then no matter how complicated the question, it can be asked by a relational database. Some queries are too complicated to work with flat files.

## Using a professionally-designed program to analyse sports injuries

Professionally written databases will become more common in the future as more researchers have a need to manage large amounts of data. A professionally-written database can be an off-the-shelf product, or can be custom written by a programmer after determination of the requirements for the research. Custom products are currently very expensive because the market is currently small. Off-the-shelf products have the disadvantage of needing to be written for a large number of users, so it is difficult to strike a balance between functionality for everyone and huge unwieldy menus of functions that the majority of users do not want. The product of the future will probably have an off-the-shelf framework that each user will have the opportunity to modify at the time of purchase, so that it works most efficiently in its environment.

Off-the-shelf programs to monitor and analyse injuries are currently available, such as Injury Tracker and Sport Care. Both these programs are based within a relational database environment. Injury Tracker (*http://www.injurytracker.com/*) is a program written within the dBase environment, whereas Sport Care (available at *http://www.humankinetics.com/*) is written within Microsoft Access, although neither of the programs require the user to have a copy of the parent database program. Both of these programs enable the user to analyse injuries across a wide range of sports, after recording injury information, clinical notes and test results. They have been designed for the North American market and are particularly suitable for athletic trainers who look after athletes from multiple teams in a school or college environment. The features of these programs are less relevant as users move further from these typical environments. A similar program is being developed for the UK market (*http://www.sportsinjurymanager.co.uk*). Med Sports Systems (*http://www.med-sports-systems.com/*), a company established in the USA by John Powell, also sells sports injury monitoring software (SIMS), written within the FoxPro parent database program. This product is often purchased by entire competitions as part of an overall injury surveillance system, where the company will not only provide the database, but also collate and report on the injury statistics. The most established client of Med Sports System is the National Football League, which has required all teams to use a standard injury database

for over 20 seasons.[5] Orchard has provided similar injury surveillance services to sporting bodies in Australia such as the Australian Football League, enabling a large injury database to be established.[6]

## Conclusion

Sports medicine researchers already use databases on regular occasions in the process of conducting their own research, writing a scientific paper, or reading the work of others. The ubiquitous presence of databases in sports medicine will only increase in the future. An understanding of how databases work, and skills in using at least one of the major databases in each of the categories reviewed in the chapter, will be mandatory for the sports medicine researcher and clinician of the future.

### Key messages

- The power of internet search engines is due to databases that relate websites to key search phrases
- Medline and SportDiscus are the most comprehensive literature databases to search in the sports medicine field
- Medline, in its PubMed version, is available free of charge on the internet
- Citation databases are used to file reference details when writing a scientific paper and to automatically format the bibliography when submitting the original or revised paper
- A relational database is the most powerful type of program to track injury records or injury-related details in a clinical setting

## Sample examination questions

### Multiple choice questions

1 A relational database is:

A A database program that is related to another program in an Office Suite
B A database where the data is stored in multiple tables that are linked by relationships between them
C A program such as a spreadsheet that is used as a database
D A program such as Lotus Notes, which can run queries on data
E Data outside a database that is related to data within a database

2 SportDiscus differs from Medline in that:

A It is available on the world wide web
B It is available free of charge
C It contains a greater number of sports medicine journals in its database
D It is more commonly used
E It does not provide abstracts of references within the database

3 Which of the following programs are citation databases?:

A Microsoft Access
B EndNote
C Lotus Notes
D Injury Tracker
E Sport Care

**Essay questions:**

1 You are the medical director of a sports medicine centre situated within a university that treats all athletes from the university sporting teams. You would like to write a scientific paper that compares the injury rates from the different types of sport that are played at the university. Describe three ways in which you could use a database to help you conduct this study and write a paper for submission to a scientific journal.
2 List some sports medicine journals that are indexed in PubMed, and some journals that are not. As a sports medicine researcher, why is it important that you have an idea which of the journals are included in PubMed?
3 For a surgeon who wants to keep a computerised record of operative details, comment on the advantages and disadvantages of using a spreadsheet compared to a relational database.

## References


1 Haynes B, McKibbon K, Walker C, Ryan N, Fitzgerald D, Ramsden M. Online access to MEDLINE in clinical settings: a study of use and usefulness. *Ann Int Med* 1990;**111**:78–84.

2 Orchard J, Hayes J. Using the world wide web to conduct epidemiological research: an example using the National Basketball Association. *Int J Sportsmed* 2001;**2**(2).

3 Scherrer C, Dorsch J. The evolving role of the librarian in evidence-based medicine. *Bull Med Libr Assoc* 1999;**87**:322–8.

4 Sable J, Carlin B, Andrews J, Sievert M. Creating local bibliographic databases: new tools for evidence-based health care. *Bull Med Libr Assoc* 2000;**88**:139–44.

5 Powell JW, Schootman M. A multivariate risk analysis of selected playing surfaces in the National Football League: 1980 to 1989. An epidemiologic study of knee injuries. *Am J Sports Med* 1992;**20**(6):686–94.

6 Orchard J, Seward H, McGivern J, Hood S. Intrinsic and Extrinsic Risk Factors for Anterior Cruciate Ligament Injury in Australian Footballers. *Am J Sports Med* 2001;**29**(2):196–200.