EVALUATING

DATABASES

Richard Wilson
Gareth Parry
Jennifer Burr

January 2005
Evaluating Databases

Richard Wilson¹, Gareth Parry¹, Jennifer Burr²

Marc Chattle, Health Services Research, School of Health and Related Research, University of Sheffield, for administrative support

On behalf of the Review Body for Interventional Procedures Programme, Health Services Research, University of Sheffield¹ and the Health Services Research Unit, University of Aberdeen. ²

Jennifer Burr
Health Services Research Unit
University of Aberdeen
Polwarth Building
Foresterhill
Aberdeen, AB25 2ZD
Tel: (01224) 559715
Fax: (01224) 554580; Email: j.m.burr@abdn.ac.uk

Marc Chattle
ReBIP Project Administrator
Health Services Research
School of Health and Related Research
University of Sheffield
Regent Court, 30 Regent Street
Sheffield, S1 4DA
Tel: +44 (0)114 2220742
Fax: + 44 (0)114 2220749; Email: m.chattle@sheffield.ac.uk

January 2005
# TABLE OF CONTENTS

1. INTRODUCTION .................................................. 3
2. OBTAINING THE INFORMATION ............................... 3
3. DEFINITIONS .................................................... 3
4. DATABASE EVALUATION ....................................... 4
   4.1 Main criteria .............................................. 4
       4.1.1 Management and organisation ...................... 4
       4.1.2 Data set and quality ................................ 5
       4.1.3 Ethics and governance .............................. 6
       4.1.4 Data analysis ........................................ 6
       4.1.5 Commercial sponsorship ............................ 7
5. SUMMARY ........................................................ 7
6. REFERENCES ..................................................... 8
7. APPENDIX ......................................................... 8

APPENDIX .......................................................... i
1 INTRODUCTION

This report was presented to the Interventional Procedures Advisory Committee (IPAC) at the National Institute for Health and Clinical Excellence on Friday 11 March 2005.

The Interventional Procedures Programme (IPP) provides guidance on the safety and efficacy of interventional procedures. Most procedures reviewed by the IPP are ‘new’ and the evidence on the safety and efficacy of the procedure can be limited. For certain procedures a database exists, usually hosted by either the relevant specialist society or the manufacturer. In these circumstances prospective data collection onto an existing database could be a recommendation in the IPP guidance.

The purpose of the report is to assist IPAC in its assessment of the quality of an existing database and whether prospective data collection should be recommended to assess the safety and efficacy of the interventional procedure.

2 OBTAINING THE INFORMATION

The database literature ought to identify the management team or data custodians. If there is insufficient information on the literature that is made available (either from a database website or other sources) then an approach will need to be made to these individuals. There is one very useful source of information that merits special attention. The “DocDat” organisation has built up an impressive directory of UK databases. Each database is assessed using a standard questionnaire (where many of the items are numerically scored thus allowing comparative histograms to be drawn) and covers the background and management of the database as well as explicit questions dealing with the data themselves. It seems sensible, therefore, that when called upon to assess any database, the first port of call ought, therefore, to be DocDat. However, not every database is necessarily going to be found within this directory.

3 DEFINITIONS

A number of terms are used in the literature. One can speak of “registers”, “registries” or “databases”. The registry is the location or organisation wherein which a register is held. Any one registry may house several registers. According to Newton and Garner a clinical
database differs from a register in that it usually has no defined denominator population. Registers by contrast will have this relationship to a denominator population and may be used for epidemiological purposes and the determination of for example prevalence or incidence rates. Within this document the expression “database” will be used throughout.

4 DATABASE EVALUATION

A database may be evaluated across a large number of criteria. Texts from Klaucke, Pryor, Black, and Newton and Garner have been scrutinised in order to give as wide a range of opinions and viewpoints as possible. Klaucke approaches evaluation from an epidemiological viewpoint and makes mention of the simplicity, flexibility and acceptability of the “surveillance system” as being important. For Pryor the characteristics of a successful database include good leadership by a multidisciplinary team, stable funding and focused goals. Newton and Garner, reiterate the points made by Klaucke while Black concentrates on factors relating to the quality of the data set.

The following section lists headings commonly used when evaluating databases. For some of the criteria a subjective assessment may have to be made of the extent to which they are being met, in others it may be possible to numerically “score” them.

4.1 Main criteria

The criteria commonly used in evaluating databases are presented and discussed here. This section can be read in conjunction with the accompanying ‘checklist’.

4.1.1 Management and organisation

This section relates to those aspects of the database dealing with its management and purpose.

Focus

The database should have a clear aim. Poorly defined aims may lead to inappropriate collection of data (irrelevant or poor quality).

Team

Effective leadership by a well organised multi-disciplinary team is considered a hallmark of a good database. Leadership refers to more than those tasked with data gathering or inputting, as it encompasses all those who are charged with the management of the database
and the determination of its trajectory. The team may include interested clinicians, epidemiologists, health service researchers, statisticians and IT professionals in order that the database in its totality can be directed and managed.

**Funding**

The database should have secure long-term funding to enable it to carry on its work for as long as is required. Databases are (or can be) expensive to run. If data processing, programming and management staff are employed to oversee the work of the database, the cost will run into tens of thousands of pounds per annum. Funding itself may come from research or other grants or there may be support from a professional body. There appears to be no intrinsic reason why purely commercial funding should compromise the value of a database, assuming all other aspects are satisfactory.

4.1.2 **Data set and quality**

The practical utility and value of the database is dependent on the security of the data and how effectively and efficiently the gathering and recording of the data is carried out.

**Security**

It is important that all relevant legislative requirements on data security and handling are complied with. This covers such matters as the storage of data (whether paper or electronic), its movement or transmission to and from the database and access to the stored data. Procedures should be in place that clearly address these issues. There are serious legal and ethical implications for a database that fails to observe these precautions and failure to tackle these issues may indicate weaknesses or poor practice elsewhere in the management of the database. Increasingly databases are opting for electronic transmission and handling of data. This does raise concerns over the potential for hacking or corrupting data (which can be protected against through the use of encryption and firewalls etc.,) set against which is the practical benefit gained in terms of convenience for those submitting data.

**Compliance**

If the database is failing to pick up cases which ought to be registered (it records a low proportion of the “relevant population”) this will degrade the quality of any analysis that is undertaken. The database management team should have procedures for checking how complete the database is, for example, through surveying of clinicians in specialist societies and comparing this with the data in the registry. To some extent this links back to comments
made under “Funding”. If a specialist or professional body recommends the database then this may assist in achieving good compliance rates.

**NOTE:** DocDat numerically score this item on four levels, with level 1 being defined as including 80% or less of the eligible population and level 4 being 97% or more of the eligible population being recorded in the database.

*Validation*

Even if all suitable cases do make it onto the database there remains the question of the extent to which the recorded information reflects the actual clinical information on which it is supposedly based. There should be a plan for validating and checking input data to ensure that what is recorded is accurate. If validation either cannot or is not being undertaken then this will cast doubt on the value of analytical output. However, validation may be expensive and time-consuming so may be seen as a less practical option for databases with limited funding.

**NOTE:** one could also consider here the completeness of the records in terms not of coverage, but the extent to which each individual case record holds data on the variables. DocDat score this criterion on four levels examining the presence, or not, of range and consistency checks and validation against external sources.

### 4.1.3 Ethics and governance

It is possible that the database will need to be approved by an ethics committee if the remit of its data gathering exceeds that information which is routinely collected for these individuals as part of their normal care. While this aspect of the database is not critical in a functional sense, there is an obligation on those who gather and handle data to ensure that they comply with all relevant UK (or European Union) requirements. In addition to the ethical clearance it is increasingly likely that permission will be needed on an individual basis from the research departments of all trusts from which information is being supplied. This is a variable process, only recently introduced, and some database custodians may be unaware of their responsibilities in this area.

### 4.1.4 Data Analysis

The managers of the database ought to have a plan of analysis and dissemination or may have already made available the results of previous analyses. Engagement with those who supply the data is important and provision of analysis at the local level can provide valuable
feedback to individual reporting clinicians. There may also be available copies of published reports based on the database contents.

### 4.1.5 Commercial sponsorship

Records may be maintained by manufacturers of the use to which their products are put and the outcomes that result. The company may provide financial or other material backing to support such a database. That there is such a link to a commercial concern may not in itself compromise the value of the database, although there appears no discussion of this in the literature. It would seem more useful to assess this type of database in the same way that a professional society database would be assessed. If the commercial database exhibits high standards in terms of its leadership, compliance, quality validation etc., then it might be considered a useful tool for the purposes of the Interventional Procedures Programme.

### 5 SUMMARY

The preceding text has described the aspects or features of a database for assessing its “quality”. While this is not an exhaustive discussion of the issues, it does, however, give an indication of what attributes might separate out the better database from the more not so good. Having said that, it needs to be remarked that a final assessment of the database’s quality or utility will depend on more than just these points. In an ideal world all databases would meet every criterion fully. In a less than ideal world the picture is much more mixed. Reaching a decision about a database will require consideration of other wider factors. The database in question may be less than ideal (according to these criteria), but may be the only database in existence for that intervention. If it is less than satisfactory it may be possible to improve it rather than attempt to establish a new database (time consuming and expensive) which will reach the required standard.
6 REFERENCES


7 APPENDIX

To assist with the practical assessment of a database, the following two pages make-up a ‘checklist’ of those items which seem the most pertinent when undertaking the evaluation. The wording and structure of the questions have been adapted from the ‘DocDat’ data collection manual published on the Directory of Clinical Databases website supplemented by some additional questions developed by the Review Body team.
APPENDIX Checklist

PART 1 - BACKGROUND

1) WHAT IS THE PROCEDURE COVERED BY THE DATABASE/REGISTER?
   ---------------
   
2) WHAT IS THE PURPOSE OF THE DATABASE/REGISTER?
   
   - to assess the safety of the intervention
   - to assess efficacy of the intervention
   - to assess effectiveness of the intervention

3) WHAT IS THE GEOGRAPHICAL AREA COVERED BY THE DATABASE/REGISTER?
   ---------------
   
4) HOW MANY INDIVIDUALS OR EPISODES OF CARE ARE INCLUDED IN THE DATABASE/REGISTER?
   ---------------
   
5) WHEN WAS THIS FIGURE DETERMINED?
   ---------------
   
6) WHAT IS THE TIME PERIOD COVERED BY THE DATABASE/REGISTER?
   ---------------
   
7) WHAT IS THE TIME FRAME THAT GOVERNS RECRUITMENT TO THE DATABASE/REGISTER?
   
   - continuous
   - periodic
   - one-off

8) DOES THE DATABASE/REGISTER TRACE INDIVIDUALS THROUGH MORE THAN ONE EPISODE OF CARE?
   
   - yes
   - no
   - unknown

PART 2 – MANAGEMENT AND ORGANISATION

1) WHO IS INVOLVED IN THE MANAGEMENT OF THE DATABASE/REGISTER?
   ---------------
   
2) IS THE DATABASE/REGISTER APPROVED BY ANY CLINICAL OR PROFESSIONAL BODIES?
   
   - yes
   - no

   If YES, BY WHOM?
   
   - private sector organisation
   - public sector organisation
   - clinical/professional body

3) WHAT IS THE SOURCE OF FUNDING?
   
   - public sector
   - private sector
   - charity
   - membership subscription
   - none

4) HAS THE DATABASE/REGISTER RECEIVED ETHICAL APPROVAL?
   
   - yes
   - no
   - not applicable

5) HAS THE DATABASE/REGISTER RECEIVED RESEARCH GOVERNANCE CLEARANCE FROM PARTICIPATING TRUSTS?
   
   - yes
   - no
   - not applicable

PART 3 - SECURITY

1) FOR A DATABASE/REGISTER THAT USES A WEB INTERFACE, DOES THE LOG-IN REQUIRE:
   
   - user name
   - user name plus password
   - user name, password and authorised location
   - none of these

2) WHERE IS A BACK-UP VERSION OF THE DATA HELD?
   
   - stand alone computer (not networked)
   - computer connected to external network
   - other electronic storage media (e.g. CD-ROM, mass storage device)
   - no back up

3) ARE PAPER FORMS (IF USED) STORED?
   
   - yes
   - no

4) ARE THERE PROCESSES IN PLACE FOR MAINTAINING THE AUDIT TRAIL?
   
   - yes
   - no

5) HOW ARE RECORDS STORED ON THE DATABASE/REGISTER?
   
   - irreversibly anonymised
   - reversibly anonymised
   - identifiable

6) HAVE SUBJECTS GIVEN CONSENT FOR DATA COLLECTION?
   
   - signed consent for each episode of care
   - signed consent only on first contact with service
   - signed consent not obtained but option to opt out
   - signed consent not obtained & no option to opt out

PART 4 - THE DATA SET

1) DOES THE DATABASE/REGISTER HAVE A WEB INTERFACE OR USE PAPER FORMS?
   
   - web interface
   - paper forms

2) IS A COPY OF THE DATA COLLECTION TOOL AVAILABLE?
   
   - yes
   - no

3) IS A LIST OF VARIABLES AVAILABLE?
   
   - yes
   - no
4) What is the primary outcome the database/register records?

5) Are the relevant safety outcomes assessed?

6) Are the relevant efficacy outcomes assessed?

7) Are nationally approved codes used for identifying:
   - subject: yes/no
   - clinician: yes/no
   - institution: yes/no

8) Is the data linked to other database data?

**Part 5 – Quality Assurance**

1) Has the relevant population been included in the database/register?

2) How complete is recruitment of the relevant population?
   - unknown of few (<80%)
   - 80-89%
   - 90-97%
   - >97%

3) How and when was completeness last determined?

4) What percentage of the continuous variables are collected as raw data?
   - <70%
   - 70-89%
   - 90-97%
   - >97%

5) What percentage of the variables have clear definitions (as defined in a data manual)?
   - none
   - <50%
   - 50-97%
   - >97%

6) What percentage of the variables have explicit rules on their coding (as defined in a data manual)?
   - none
   - <50%
   - 50-97%
   - >97%

7) How reliable is the coding (as determined by Kappa scores)?
   - not tested
   - Kappa <0.5
   - Kappa 0.5-0.8
   - Kappa >0.8

8) Are range checks built into the database/register design?
   - yes
   - no

9) Are consistency checks built into the database/register design?
   - yes
   - no

10) Is the recorded data validated against case records?
    - yes
    - no

**Part 6 – Analysis**

1) How frequently are data transferred from health care providers to the central database/register? (relevant for paper-based databases)

2) Can individual clinicians access their own data?
   - locally: yes/no
   - centrally: yes/no

**Part 7 – Publications**

1) How frequently are multi-centre audit reports produced?

2) How frequently are provider-specific reports produced?

3) Is there a bibliography of published papers and/or reports based on data from the database/register?