Data quality and clinical audit

Ranjit Verma

Abstract

Clinical audit has an increasingly important role in the quality of care being offered to patients. Only good-quality data can enable valid conclusions to be drawn, which in turn enable changes to be made for the better. More and more patient data are held on NHS computer systems, and increasingly these data are being used to facilitate the audit process. The quality of any such data needs to meet the highest standards. This article briefly defines the audit cycle and goes on to consider a typical data model. The various elements of the data model are defined, the understanding of which should enable individuals to avoid pitfalls in data collection and ensure that the data they collect for clinical audit are of the highest quality.

Keywords audit; clinical audit; data quality; informatics; quality assurance

The practice of clinical audit is now well established within the healthcare profession in the UK, and is a requirement for most NHS job contracts, including those for trainees. Clinical audit helps to improve the quality of care delivered to patients and is invaluable in maintaining and monitoring standards of care. Clinical audit acts as a motivating factor for individuals by sharing outcomes and identifying areas of concern, so that timely remedial action may be taken. Its success partly relates to the fact that individuals have full control to drive the whole process with support from colleagues and institutions.

The audit cycle

Clinical audit is a cyclical process, where standards are agreed and data collected. Findings from the analysis of collected data show whether or not standards are being met. If they are not being met, changes are planned and implemented, and data are collected for a second time and analysed to see if any improvements have resulted from these changes (Figure 1). It is important to realize that data are collected and analysed on two occasions. A single data collection exercise does not constitute an audit. The first data collection is done to establish the current position, and the second is done to see if any improvements have been made (see *Anaesthesia and Intensive Care Medicine* **5:** 410–2).

Ranjit Verma *FRCA is Consultant Anaesthetist at the Derby City General Hospital. He qualified from the University of Manchester and is the immediate past Chairman of the Society for Computing and Technology in Anaesthesia (SCATA).*

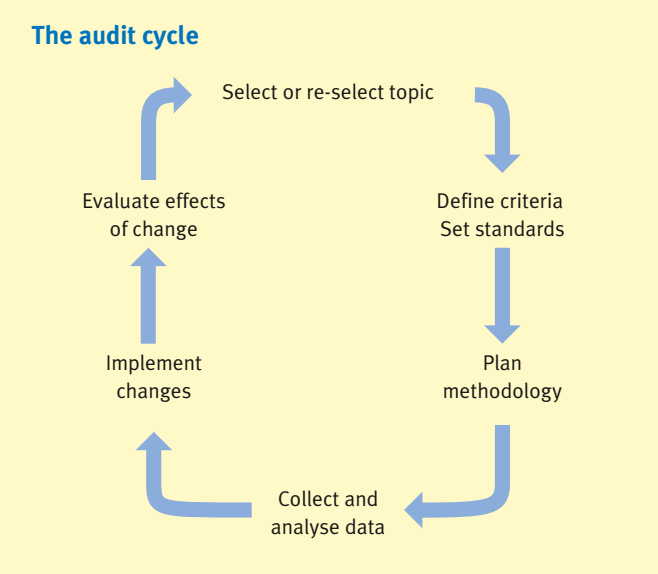

Figure 1

Raw data need to be assimilated and processed before being turned into useful information. The quality of this information will depend on the quality of the data collected.

As data collection is an integral part of the audit cycle, any system that relies on collecting and turning data into useful information will fall at the first hurdle if the quality of data collected is poor. Particular care must therefore be taken to ensure that the data collected are good enough to enable valid conclusions to be made.

A data model

In determining the type of data that should be collected for a given audit project, due consideration should be given to the factors below.

The application: why are the data being collected? There should be a clear understanding of the purpose for which the data are being collected, and it should be clearly stated.

Data collection: a description of the exact process by which the data will be collected should be stated in the audit. Will the data collection be retrospective or prospective? Will data be collected on paper or directly onto an electronic format? Very often data are first collected on paper and then transferred onto an electronic database.

Data warehousing: this is a term to describe the systems and processes that will be used to store the data. Usually this is some form of electronic storage (e.g. a database or spreadsheet). There may sometimes be elements of some intelligent processing or calculations of the raw data at this stage.

Data analysis: the bulk of logical processing takes place at this stage. The raw data are transformed into useful information. Unlike research, data collected for clinical audit are not usually subject to the rigours of statistical analysis. Simple percentages

and review of absolute numbers is usually acceptable. Where possible, these figures should be compared with audits performed by peer groups or national standards.

After careful planning a data set is identified. In order to ensure that the raw data are as viable as possible, each element of the data set should be considered against the criteria discussed below.

Accuracy: are the data elements accurate and are the values correct and valid? A data value can be correct and valid at a given time point, but may not necessarily be so later. For example, a population or demographic statistic can vary during a period of time. An individual blood pressure reading or a blood glucose level may be relevant and significant at a given time point, but may have no relevance at another.

Accessibility: are the data easily accessible from their source? This often depends on how data have been stored. Prospective data are usually planned out and kept in such a way that they are relatively easy to access. Retrospective data may not be as easy to access. Often the data are stored on sophisticated computerized systems, where the end-user may have little or no control over access. Security and confidentiality are serious issues. Accessing data from hospital information systems may require authorization from senior management, and a formal application stating exactly why the data are required and how they will be used may need to be submitted. The data collected by an NHS Trust about an individual are the property of that Trust, although the individual has a right, under the Data Protection Act, to have complete access to their data (see Office of Public Information, formerly Her Majesty's Stationery Office, website http://www.opsi.gov.uk/ACTS/acts1998/19980029.htm). This applies only if the individual is identifiable from the data set. If the data set does not contain any identifiers then the Data Protection Act does not apply. Once permission has been obtained, accessing the data may take several weeks or even months, which may slow down the process significantly.

Comprehensiveness relates to how much data should be collected. There is a tendency to collect far more data than necessary for a given project. This is counter-productive and a waste of resources. Further, unnecessary data make a study more complicated and can slow down the system at every level (e.g. during data collection, storage and analysis). Only data that are absolutely necessary for the audit project should be collected. Occasionally it is not possible to collect all the data that are necessary. In such cases, whether or not the audit project will be viable has to be decided. If it is deemed necessary to proceed with the project in the absence of a complete data set, the limitation should be clearly stated.

Consistency reflects the reliability of the data items. If the data are not reliable and reproducible, the audit project may be considered to be spurious or a one-off, and will lose its importance and significance.

Currency: the data should be as up-to-date as possible. For some audit projects the data may not have been current at the time of collection, but may have been current for a time point specific to the audit project. In such cases, the data may be good enough to use, providing the project concerns that specific time point. This is often the case for projects looking at retrospective data. For example, when using retrospective data, a patient's age can never be the same as their current age.

Definition: each data item should be defined clearly. This ensures that current and future users know exactly what the data elements mean. Each data item should have a clear meaning and the values attributed should be acceptable. For example 'out of hours' is a frequently used term to describe work done during unsocial hours. However, there is no universal agreement as to the definition of this term. Does it relate to work done between 9.00 p.m. and 9.00 a.m., between 10.00 p.m. and 8.00 a.m., between midnight and 9.00 a.m. or some other period of time? If this term is to be used it should be clearly defined.

Granularity means that the attributes and values of each data item should be defined at the correct level of detail. For example, the 'sex' field of a database is usually configured to accept an input of 'male' or 'female' only. However, what about mixed gender issues (e.g. hermaphrodites)? What about gender-change operations, where a male or female patient changes sex at completion of the operation? To prevent such inconsistencies the data input should take account of such details. Hence, for the above example the 'sex' field should allow five inputs reflecting male, female, hermaphrodite, male to female and female to male!

Precision: data values should be just large enough to support the application or process. If the data values are too small the data set may be incomplete or erroneous. Data values that are too large can potentially give rise to the input of erroneous data or the collection of unnecessary information. For example, at first sight a two-digit 'age' field in a database may be considered adequate (i.e. the age range between 1 and 99 years). This is fine as long as you are not expecting any patients in your sample to be over 99 years of age. Although it is uncommon, it is certainly possible to have a patient aged 100 years or more. At the opposite end of the age spectrum children may be under the age of 1 year or 1 month. Hence, in determining a patient's age, a 'date of birth' field may be preferable, where age can be calculated by working out the difference between the present date and the date of birth.

Relevancy: the data should be meaningful for the purpose of the project. There is no point in collecting unnecessary data; it is not only time consuming, but also can slow the whole process to a grinding halt.

Timeliness will depend on how the data are to be used. During a long and complicated operation instantaneous blood pressure readings may need to be displayed at all times. Hence, the monitoring systems should be able to capture, process and display the data on the monitor screen on a beat-to-beat basis in real time. However, data regarding numbers of cases completed in a given month are required as a monthly statistic only, so that future services can be planned according to the findings.

Data in clinical practice

Data in clinical practice produce two types of information: primary and secondary.

Primary information is used for immediate management of the patient's health (e.g. data for diagnosis or for planning treatment such as blood test results, radiographs, etc.).

Secondary information is used for aspects of care that are not immediate, such as planning services, performance management, adverse incidents, clinical governance, risk management issues, etc.

Data input errors

There is no system that is free from data-capturing errors as they are caused by many factors. Direct capture of data from the source to an electronic device without interference by humans is the best way to minimize data input errors. However, when human interface is unavoidable input errors may occur and consequently degrade data quality. Some input errors are described below.

Missing information: incomplete data sets are common, especially when data are collected retrospectively.

Typographical errors: if data relies on humans transferring data from one format to another, data errors are inevitable. For example, if data are collected on paper and then typed into an electronic database (the most frequently used method of capturing data in clinical audit), typographical errors are likely to occur. Tedium, monotony, boredom and lapses of concentration all play a part. Error can sometimes be reduced with multiple-choice answersheets, where responses are limited to tick-boxes and no free text is permitted. Data capture can be further improved by taking the computer keyboard out of the equation and using bar codes, optical mark readers or optical character recognition systems.

Interpretation errors: clinical terms are often subject to differing interpretations. To minimize this type of error, data should be coded. The code, with its specific meaning, is contained in a data dictionary. Coded information is easier to store and manipulate electronically.

Coding systems

A full discussion of coding of clinical terms is beyond the scope of this article. However, there are several coding systems in use by the NHS information systems, which are constantly being refined and developed. They are designed to enable clinical information to be stored electronically and to allow healthcare professionals direct access to this information regardless of geographical location. Examples of these systems are given below.

Office of populations, censuses and surveys version 4 (OPCS-4) is a coding system that classifies surgical operations and procedures. This classification has been devised for translating or classifying all surgical procedures that may be carried out during healthcare episodes. The most recent version (4.3) was introduced in April 2006.

International classification of diseases version 10 (ICD-10): this system codes diseases and health-related problems.

SNOMED and SNOMED-CT: this systematized nomenclature for medicine was developed as a thesaurus of approved clinical terms and their synonyms. It uses a unique tree structure to code clinical terms – http://www.snomed.org/snomedct/

International classification of functioning, disability and health (ICF) is a further coding system being developed in collaboration with 65 countries and will complement existing coding systems.

Summary

Clinical audit has an increasingly important role in the quality of care being offered to patients in the NHS. Only good-quality data can enable valid conclusions to be drawn, which in turn enable changes to be made for the better. More and more patient data are held on NHS computer systems and increasingly these data are being used to facilitate the audit process. The quality of any such data needs to be of the highest standard. ◆