

VIEWPOINT

# Rules for collection of clinical data: the eight Rs

Nigel Zoltie

**Abstract**

Many reports in medical journals give advice on statistical analysis of data collected in clinical studies. The terms used are statistical, and clinicians often ponder how their data stand up to these analyses. Eight simple, easy to remember words are suggested that will help clinicians both to interpret their own data and to assess the accuracy of presented or published data.

(J Accid Emerg Med 1996;13:127-128)

Key terms: data collection; statistics; assessment

Data collection has become increasingly important in medical care, whether for audit, research, finance, planning, or a multitude of other reasons. When setting out to collect data or when attempting a critical analysis of results based on other presented data, some guidelines are often helpful for the practising clinician. These guidelines may not be as precise or as far ranging as formal statistical analysis, but allow a quick appreciation using simple memorable concepts that may form the basis for further analysis. Eight attributes seem to cover most requirements for good quality data (table): these attributes (the eight Rs) can in principle be applied to almost any form of data collection.

**(1) Data should be recorded**

OED: Recorded – set down for remembrance or reference, put in writing or other permanent form.

This is the first rule of data collection, and may seem self evident. A moment’s reflection, however, will remind us that in the clinical setting this omission of data recording is a very common failing: the detailed history and examination that is often obtained but is recorded as a two line summary on the accident and emergency (A&E) card is a prime example. How often, in the resuscitation room, is activity performed but scantily recorded (and usually only in retrospect)? This problem of failure to record data is at its most significant in retrospective clinical studies. In particular it is impossible to tell with certainty whether data that have not been recorded were collected but

not recorded, or not collected. Prospective studies can avoid the problem by defining before the study how to record the data, and by making the data recording simpler, for example by the use of forms or tick boxes. But for retrospective studies the fact that data may have been collected but not recorded must be appreciated and steps taken if at all possible to estimate any effect this lack of recording might have on the final analysis.

**(2) Data should be readable**

OED: Readable – able to be read.

There is no point in recording data that are not readable. This term “readable” may refer to both humans and machines. It may mean that clinical notes must be legible, so that other humans can read and collect the data. It may mean that the data should be machine readable, such as forms for use in optical mark readers. These two different requirements for readability are often in conflict: what is legible and assimilable by humans is often not readable or assimilable by machines, and vice versa. Careful consideration must be given to the structure of data to be collected according to the method of reading the data.

**(3) Data should be representative**

OED: Representative – typical of a class, containing typical specimens of all or many classes.

For data to be applicable to other situations, the study sample must be representative. A clinical example might be collection of data from patients presenting to A&E departments with abdominal pain: such data might not apply to patients with abdominal pain on the wards, where the disease spectrum is different, or to another hospital where surgical patients are admitted direct to the wards, with only unusual cases attending the A&E department. So if one wishes to extrapolate from the findings in one circumstance to other areas, then one must be sure the data are representative of that new situation.

**(4) Data should be right**

OED: Right – correct, true.

Data are always correct at source. However, there are many actual steps between the source data and its recording for analysis. A clinical example might be the pulse rate: measured by ECG the heart rate may be recorded as 147, but measured by the pulse oximeter as 131. Measured at the radial artery over 10 seconds and multiplied by 6 the rate might be 126. Which is the “right” rate? And if a study records the pulse, which method has been used?

*The eight rules of data collection*

- 1 Recorded
- 2 Readable
- 3 Representative
- 4 Right
- 5 Reliable
- 6 Reproducible
- 7 Revelant
- 8 Restricted

Accident and  
Emergency  
Department, Leeds  
General Infirmary,  
Great George Street,  
Leeds LS1 3EX  
N Zoltie  
Correspondence to:  
Nigel Zoltie FRCS.

**(5) Data should be reliable**

OED: *Reliable* – that may be relied upon.

Reliability is a reminder about the need for data to be consistent. Often reliability is lacking because of missing data, inherent inconsistency in the data, or inconsistency in collection of the data. An example might be measurement of respiratory rate in acute trauma. Often the data are missing; or the rate may vary over time, and thus measurement at one point in time may not reliably indicate rate over a different time span; or measurement may be haphazard, guessed, or approximated rather than measured.

**(6) Data should be reproducible**

OED: *Reproducible* – able to be copied.

Data can only be reliable if the same source can provide the same data each time, no matter who is recording them or when. If different observers collect the same data and get different results, the data are not reproducible and hence reliability suffers.

**(7) Data should be relevant**

OED: *Relevant* – bearing on or having relevance to the matter in hand.

There is currently a tendency to collect large amounts of data – since with computers the handling of data often does not present a problem – and then to trawl these data for relevant factors in the question under study. While occasionally this approach does have merit, on most occasions it heavily increases the effort of data collection and input, with very little extra gain in interpretation of the result. Furthermore, many classic examples of supposedly relevant data have later been shown to be irrelevant. For instance, the measurement of temperature and pulse rate in acute appendicitis used to be considered important but is now known to be irrelevant to the diagnosis, which is made by other symptoms and signs. There is thus no need to collect these data.

**(8) Data should be restricted**

OED: *Restricted* – confined or limited.

The human ability to assimilate information is limited, and too much data are well known

to confuse rather than help. This concept of “fogging” may not always apply to computers, which are able to process large amounts of data rapidly. However, the conclusions they may come to are not always relevant to the question asked; indeed the first law of informatics has been proposed, that data and their interpretation are restricted to the information relevant to the specific question alone.

**Discussion**

These eight “Rs” represent a simple mnemonic-type method for assessment of data collection. They may be used as rules of thumb before setting out on a study, or for assessing the results of a completed study. They do not supersede careful statistical input to a prospective study, or careful analysis of a completed one. They may give rapid suggestions, however, as to where closer analysis is justified.

Methods do exist for reducing errors in data collection, once the need has been recognised. Guidelines, definitions, teaching of standard rules, and agreed standards all help reduce errors in reliability and reproducibility. Good study design and continuous internal audit of data collection reduce errors in representativeness and confirmation that the data collected is right. Pilot studies may help delineate which data are relevant to the question under study, and the eagerness to collect all possible data must be restrained to ensure restriction of data to only those that are relevant.

It is immediately apparent that most of the above methods are already incorporated into many good quality clinical studies, but it does no harm to remember the basics of data collection and subject the collection process to as much scrutiny as the end results of the study are usually given. Good study design should give equal consideration to the methods of data collection and to their later analysis. By this means the data collected can be used effectively and efficiently, and the best use made of such data, an aim that we would all surely endorse.