RESEARCH SERIES

Data collection in the emergency setting

M O Columb, P Haji-Michael, P Nightingale

Emerg Med J 2003;**20**:459–463

This is the ninth article in the research series and focuses on the collection of data.

When discussing data collection in the emergency setting, we first need to answer two questions; what are we trying to measure, and which group or population is to be studied? Recent advances in information technology have seen the generation of a veritable sea of data. This includes demographic and administrative information, and the routine collection of audit and activity data now deemed compulsory by some regulatory authorities.¹ Plus, there is the mass of digital information effortlessly generated by current patient physiological monitoring systems. Despite, or even because of, this apparent glut, data are often of poor quality, and may not be accurate or valid. Subsequent research is then more likely to be open to sample bias, methodological irregularity, and a lack of statistical power.²

This is important because there has never been a greater demand for objective, well conducted research in the fields of prehospital, emergency, and intensive care medicine. The degree and intensity of the workload in emergency medicine has increased considerably, and there have been dramatic advances in not only the range, but also the intricacy of treatments available. To answer questions on the outcome of these treatments the data collected have to fulfil certain criteria in terms of validity, completeness, reliability, and relevance. This is as true of retrospective and prospective studies as it is of audit.

Some of these issues can be addressed in terms of questionnaire and study design. However, others succeed or fail because of the conduct of the research protocol itself.

THE IDEAL DATASET

The quality of any data recorded in emergency medicine may be described by four terms (box 1).

Validity

Validity is a statistical term, which refers to how well the measure used truly assesses the characteristic it is intended to study. This is sometimes referred to as accuracy or external consistency. Ideally such a measure should have been

validated in other population groups with the same disease process and should be reproducible in terms of its predictive or scoring value. This is particularly important where the measure is of a relatively subjective variable, such as pain or general health, and for which many different scoring systems may exist.³ Resolving such problems at the time of designing and setting up a study are the only ways to improve data validity. Post hoc reanalysis of data already collected is usually unrewarding.

Completeness

It is self evident that if data are not complete then the statistical power of any study will be reduced. Bias may be induced because the patients for whom data are incomplete may be significantly different from those on whom a full set of data is collected. For instance, such patients may have died, or transferred to other centres because of greater severity of illness. Equally, in terms of pain scoring, a patient may be in such severe pain that a pain score is unobtainable and they are unable to cooperate with the study.

The commonest problem in the quality of data collected during studies in emergency medicine is that of missing data. Having specifically trained, non-clinical patient enrollers and data monitors can have a significant effect on both data quality and the number of studies performed.⁴ Although from the UK perspective such a notion carries immediate funding implications, improvements have also been seen after audit and feedback of the quality of data collected by emergency personnel.⁵

Reliability

Reliability refers to consistency with repeated trials. From a statistical point of view, reliability is the extent to which differences in measurements are attributable to random variability in the testing method, rather than to actual differences in the variable being studied.⁶ Recording, particularly of physiological variables, is highly prone to observer interpretation and modification. Data that are collected by unblinded observers may be biased toward the appropriateness and effectiveness of any novel treatment. Therefore there is an increasing shift towards either automated data collection of physiological variables, or observation and recording of such data by an independent, trained observer not involved in the clinical process.

Relevancy

It is all well and good to collect data at great length, expense, and effort but the most important aspect is often not the actual information collected but the interpretation and use it is put

See end of article for authors' affiliations

Correspondence to:
Dr P Nightingale, Intensive
Care Unit, Wythenshawe
Hospital, Manchester
M23 9LT, UK;
pnightingale@ntlworld.com

Box 1 Characteristics of the ideal dataset

- Valid
- Reliable
- Complete
- Relevant

Box 2 Influences on data quality

- Questionnaire and study design
- The emergency environment personnel
- The conduct of research

to. The biggest flaw with many studies has been the lack of focus on patient outcome as the primary end point.⁷ Data collection can only be effective if the information produced is relevant to an important clinical or service issue. It is then even more important that the questions answered by such data are actually used to implement change.

DATA QUALITY

A number of factors may influence the quality of data collected (box 2). Some are more open to correction than others, but they remain the only means to improve the quality of data collection.

Questionnaire design

Fundamentally data need to be collected with a minimum of extra effort. Careful planning, and good questionnaire design, can increase both the reliability and completeness of data collected. In terms of length, the shorter the better. The specific questions asked, and the method of scoring used, should be valid, reproducible, and repeatable. Data need to be in the format of well characterised datasets, for example, pain scoring with the visual analogue score.³ In particular, studies involving repeated measures should use scoring systems or data collections that are as brief as possible, for example, the SOFA⁸ score for sequential measurement of organ failure.

Whether a questionnaire is given to a patient or a member of the public, the underlying design issues are the same. There has to be clear information as to what data are needed; it has to be clearly laid out and, ideally, brief. The questions must focus on the questions being asked by the investigators, and it should be capable of being completed with the minimum of inconvenience. In such situations, the free provision of a pen or pencil as well as the careful use of timing (for example, during a lull in proceedings) may reap rewards. The challenge is to create a form or questionnaire that is easy to use while at the same time collecting a complete and appropriate set of data for later statistical analysis.

In particular, when designing a questionnaire to assess subjective data, you should be very aware that the way in which a question is phrased or structured could have a direct effect on the answer elicited. Answers should ideally be in a yes/no format rather than requiring free comment, or even a range of answers, as this tends to make further analysis almost impossible. The role of the structured interview in emergency medicine environments is of extremely limited value, not least because of the bias induced in any questions by the simple presence of the investigator doing the asking. Answers are always clouded by the interviewees' perspective of whether they are talking to a member of staff, an independent observer or what other role the interviewer may have, in this threatening and alien environment.

Obviously, in clinical studies on patients, the initial dataset should include documentation of informed consent. Issues such as whether a questionnaire is directly entered into a computer database, collected on paper, or whether a computer reads it, are in many ways superfluous to these initial design problems.

The emergency environment

There are many aspects of the emergency environment that militate against the conduct of good quality research. The workload is often intermittent, and may be unpredictable.

Patient casemix is also heterogeneous unlike, for example, elective surgery. Patients often have a wide variety of concomitant problems and severity of initial diagnosis. Characteristically such patients also present in a non-uniform pattern throughout the day, and often outside normal "office" hours.⁹

The emergency environment poses many barriers and obstacles to patient recruitment and data collection, and this has implications particularly for the staffing of prospective trials. The initiation of a treatment may be time limited and lead to failure to enrol potential research subjects. For example, many recent studies of sepsis often have a narrow (six hour) time period in which to identify patients, gain consent, administer the trial drug, and record the initial set of data.

Personnel

The personnel involved in the conduct of research also have an important impact on the quality of data collection. Clinical staff in the fields of prehospital, emergency medicine, and intensive care are personnel with training predominantly focused on patient care. They may have a varying degree of interest, and indeed ownership, of any research protocol. The filling out of forms, collecting of data, and entry of patients into studies may be seen as a comparatively unimportant and even superfluous issue in the heat of emergency medical treatment.¹⁰

There is evidence that specific training for staff, and regular audit of data quality may have a positive effect on the completeness and accuracy of data generated.⁵ An even better option would appear to be the appointment and training of specific and independent research enrollers and data collection staff.^{4 11}

The conduct of research

The way in which a questionnaire design is implemented also has an impact into the success of the project. An independent observer, trained to pick out the relevant data at the time of any particular procedure, is by far the most accurate and most complete way of collecting such data.^{5 9} Retrospective collection of data has been shown to be less accurate. We know, for example, that the more severe derangements of physiological variables are often minimised in the patient's records. This is one of the reasons why prospective studies are to be preferred over retrospective data collection to answer important clinical research questions.

DATA VALIDATION

Computer hardware and software have to a large extent removed the tedium from data processing. It is, however, important for the researcher to be aware of possible sources of error and to appreciate how these can be minimised. In addition, the researcher must have some understanding of the principles and assumptions behind the analyses performed lest powerful statistical techniques be misused, for example, analysing a number of repeated measurements.¹²

Precision

Data should be recorded to the precision of the measurement system. For presentation it is reasonable to report the arithmetic mean to one decimal place beyond the precision of the data. The index of variability should be reported to two further decimal places than the original data.¹³ Full precision should be retained however during analysis of the data.

Errors of transcription

Errors in data transfer can be reduced by minimising the number of times data have to be entered. As in most instances analyses will be carried out on a computer, it is preferable to use the computer screen for the initial data entry. The computerised form can be designed to carry out logical checks (see later), which should improve the accuracy of data recording.

Figure 1 Frequency histograms describing severity of left and right carotid occlusion (% stenosis). The plot shows, in paired right and left common carotid arteries, in patients referred for angiography, that the distributions are different for both vessels. Further analysis showed that the left side was more significantly diseased. (Data obtained with permission from Mr G Griffiths, Vascular Surgery, Ninewell's Hospital, Dundee.)

Smaller data files can be simply checked by two persons, one reading the original data and the other verifying the entry. For larger datasets the technique of "double entry" should be used. This entails entering the data on two separate occasions as duplicate files. Summary statistics of columns and rows can be used to check for discrepancies, or a software program can be used to compare the files.

Logical checks

Implausible or impossible data can be screened for with computerised forms. Logical checks may be set by range or category to only admit certain patients.¹⁴ For example, in a study of patients with acute lung injury logical checks might be set to only admit data from subjects whose A-a gradient or $\text{PaO}_2/\text{FiO}_2$ ratio exceeded a particular value.

Data screening

Once the data have been entered it is then useful to perform a preliminary examination of the data. This will be useful for identifying missing values and help identify those that can be retrieved from the original observations. Before deciding how to handle such missing values, it is important to consider if some systematic bias or censoring has been operating, for

Figure 2 Frequency histograms of thromboelastograph (TEG) variables. A bimodal distribution can be seen for r , k , and MA , while α is positively skewed. Further analysis of these data showed that there was a significant effect of sex and that the bimodal profile was attributable to presence of two distributions. (Data obtained with permission from Drs H Gorton and G Lyons, Obstetric Anaesthesia, St James' University Hospital, Leeds).

Figure 3 Box-whisker plots of the TEG variables. The extent of the central box denotes the interquartile range (25th to 75th centiles) with the median (50th centile) within. The extent of the whiskers represents the interval from the 2.5 to 97.5 centiles of values. Outliers beyond this range can be checked for validity.

example, loss of values because of the magnitude of the variable simply exceeding the range of the measuring system. Clustering of missing values may invalidate some or all of the effects being tested or estimated, and may indicate some bias in sampling. If the missing data are not frequent and appear random, then simple techniques such as pairwise or listwise deletion may be appropriate. Pairwise deletion results in the corresponding variable entry, to that missing, being dropped. Listwise deletion results in the removal of all observations on a subject who has one or more missing values. Considerable amounts of data may be lost however, particularly with the second approach.

Another approach is to use statistical techniques to replace missing values, particularly for balanced analysis of variance designs. At perhaps its simplest, this might entail replacing the missing observation with the mean value for that particular combination of factors, and then subtracting a degree of freedom for each factor from the final analysis. The underestimation of the variability caused by replacing with the mean value, with resultant increased power, is then offset by the loss of degrees of freedom. There are however a number of methods to estimate missing values for various types of designs based on formulaic and iterative solutions and the reader is encouraged to read specific statistical texts^{15 16} and obtain expert statistical advice.

Outliers and transformations

Data validation includes univariate analyses, where each individual variable is examined graphically and by summary statistics. It is important to verify that the data conform to the

Stem	Leaves
	69
1	024
	66667888899
2	1112244
	5677999
3	0001122333344
	5567
4	113

Figure 4 Stem-leaf plot of k . The stem-leaf plot is a simple technique that retains the original data while allowing the distribution to be examined. The data, in ascending value, are put into "bins" (represented by the [] symbol) on the right of the diagram. In this example the first bin contains the numbers 1–9, the second bin contains the numbers 10–15, the third bin contains the numbers 16–19, and so on. Under "stems" the start of each new decade is indicated. Under "leaves" the decade is omitted for clarity (for example, 1[] 024 represents the three numbers 10, 12, and 14. In this example the stem-leaf plot shows a bimodal distribution for k .

Figure 5 The value of two of the TEG variables are shown plotted against the normalised standard deviate. Each value is plotted against its relevant multiple of the standard deviation. The points should lie along the straight line in a perfect Gaussian distribution. The points for k appear to follow the line quite well. (One should note that there appears to be a systematic variation about the line, which is attributable to the bimodal nature of the distribution). It is clear for α than the distribution is far from Gaussian.

assumptions underlying the statistical analyses to be used (for example, Gaussian distribution or equality of variance for parametric analyses).¹⁷ This may also identify possible outliers. Outliers are extreme observations that appear inconsistent with the rest of the data. Graphical methods such as frequency histograms, “stem-leaf”, box-whisker, and normal probability plots, are useful in exploring the distributions of the data and to identify extreme values.¹⁸ These techniques, and specific analyses such as Grubbs’ test, only help in the identification of outliers. These data must then be verified against the original recording or measurement.

Some examples in the use of these techniques, taken from the data checking procedures from two studies, are shown in

Table 1 Some useful transformations

Types of data	Transformations
Positive skew, increasing variance	Logarithmic = $\ln x$, $\log_{10} x$
Negative skew, decreasing variance	Square power = x^2
Count; Poisson distribution	Square root = $\sqrt{x}$
Time; survival data	Inverse = $1/x$
Proportion; p , binomial distribution	Logit = $\ln(p/1-p)$; Arcsin = $\text{Arcsin}\sqrt{p}$; Probit = cumulative normal deviate

Figure 6 Logarithmic transform of α to a log-normal distribution.

figures 1–5. Unless there is strong evidence to suggest that the values are unreliable, then in general such values should be retained in the analyses. The decision to omit values should be taken before the main inferential analyses.

Another approach is the use of transformations. For data that are positively skewed (such as α) a logarithmic transform brings the data closer to a Gaussian distribution. This is sometimes referred to as the log-normal distribution. Figure 6 shows the α data after the transform. The logarithmic transform is also useful for stabilising variance when this increases with increasing magnitude of the variable. For the opposite scenario of negative skew, or where variance decreases with increasing magnitude of the variable, the square power transformation may be useful. Examples of various useful transforms are given in table 1.¹⁹

CONCLUSION

The emergency environment poses many obstacles to the conduct of good quality research and the collection of appropriate data. Many of these can be circumvented by careful design and planning beforehand. Both dedicated personnel and data audit have a role in maintaining data completeness. There are a number of statistical tests and techniques to both monitor and improve data quality and it is important that clinicians have at least some understanding of these.

Authors’ affiliations

M O Columb, P Haji-Michael, P Nightingale, Intensive Care Unit, Wythenshaw Hospital, Manchester, UK

REFERENCES

- 1 DEEDS Writing Committee. Data elements for emergency department systems, release 1.0 (DEEDS): a summary report. *Ann Emerg Med* 1998;**31**:264–73.
- 2 Singer AJ, Homan CS, Stark MJ, et al. Comparison of types of research articles published in emergency medicine and non-emergency medicine journals. *Acad Emerg Med* 1997;**4**:1153–8.
- 3 Ho K, Spence J, Murphy MF. Review of pain-measurement tools. *Ann Emerg Med* 1996;**27**:427–32.
- 4 Cabaugh DJ, Spillane LL, Schneider SM. Research subject enroller program: a key to successful emergency medicine research. *Acad Emerg Med* 1997;**4**:231–3.
- 5 Spaite DW, Hanlon T, Criss EA, et al. Prehospital data entry compliance by paramedics after institution of a comprehensive EMS data collection tool. *Ann Emerg Med* 1990;**19**:1270–3.
- 6 Karras D. Statistical methodology: II. Reliability and variability assessment in study design, part A. *Acad Emerg Med* 1997;**4**:64–71.
- 7 Jones SE, Brenneis AT. Study design in prehospital trauma advanced life support-basic life support research: a critical review. *Ann Emerg Med* 1991;**20**:857–60.
- 8 Vincent J-L, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996;**22**:707–10.
- 9 Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med* 1994;**23**:299–306.

- 10 Coats TJ. *Early changes in human physiology following injury*. [MD thesis]. London: University of London, 1998:37–42.
- 11 Roy A, Looney GL, Anderson GV. Prospective vs retrospective data for evaluating emergency care: a research methodology. *Journal of the American College of Emergency Physicians* 1979;**8**:142–6.
- 12 Columb MO. One-sided P values, repeated measures ANOVA; two sides of the same story! *Anesth Analg* 1997;**84**:701–2.
- 13 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991:487.
- 14 Armitage P, Berry G. *Statistical methods in medical research*. 3rd edn. Oxford: Blackwell Scientific, 1994:400.
- 15 Armitage P, Berry G. *Statistical methods in medical research*. 3rd edn. Oxford: Blackwell Scientific, 1994:277–9.
- 16 SPSS. *SPSS advanced statistics 6.1*. Chicago: SPSS, 1994.
- 17 Polley LS, Columb MO, Wagner DS, et al. Dose dependent reduction of the minimum local analgesic concentration (MLAC) of bupivacaine by sufentanil for epidural analgesia in labor. *Anesthesiology* 1998;**89**:626–32.
- 18 Bland M. *An introduction to medical statistics*. Oxford: Oxford University Press, 1987:60–1.
- 19 Instat. *Instant biostatistics*. San Diego: CA GraphPad Software, 1993:50.

Clinical Evidence—Call for contributors

Clinical Evidence is a regularly updated evidence based journal available worldwide both as a paper version and on the internet. *Clinical Evidence* needs to recruit a number of new contributors. Contributors are health care professionals or epidemiologists with experience in evidence based medicine and the ability to write in a concise and structured way.

Currently, we are interested in finding contributors with an interest in the following clinical areas:

Altitude sickness; Autism; Basal cell carcinoma; Breast feeding; Carbon monoxide poisoning; Cervical cancer; Cystic fibrosis; Ectopic pregnancy; Grief/bereavement; Halitosis; Hodgkins disease; Infectious mononucleosis (glandular fever); Kidney stones; Malignant melanoma (metastatic); Mesothelioma; Myeloma; Ovarian cyst; Pancreatitis (acute); Pancreatitis (chronic); Polymyalgia rheumatica; Post-partum haemorrhage; Pulmonary embolism; Recurrent miscarriage; Repetitive strain injury; Scoliosis; Seasonal affective disorder; Squint; Systemic lupus erythematosus; Testicular cancer; Varicocele; Viral meningitis; Vitiligo

However, we are always looking for others, so do not let this list discourage you.

Being a contributor involves:

- Appraising the results of literature searches (performed by our Information Specialists) to identify high quality evidence for inclusion in the journal.
- Writing to a highly structured template (about 2000–3000 words), using evidence from selected studies, within 6–8 weeks of receiving the literature search results.
- Working with *Clinical Evidence* Editors to ensure that the text meets rigorous epidemiological and style standards.
- Updating the text every eight months to incorporate new evidence.
- Expanding the topic to include new questions once every 12–18 months.

If you would like to become a contributor for *Clinical Evidence* or require more information about what this involves please send your contact details and a copy of your CV, clearly stating the clinical area you are interested in, to Claire Folkes (cfolkes@bmjgroup.com).

Call for peer reviewers

Clinical Evidence also needs to recruit a number of new peer reviewers specifically with an interest in the clinical areas stated above, and also others related to general practice. Peer reviewers are health care professionals or epidemiologists with experience in evidence based medicine. As a peer reviewer you would be asked for your views on the clinical relevance, validity, and accessibility of specific topics within the journal, and their usefulness to the intended audience (international generalists and health care professionals, possibly with limited statistical knowledge). Topics are usually 2000–3000 words in length and we would ask you to review between 2–5 topics per year. The peer review process takes place throughout the year, and our turnaround time for each review is ideally 10–14 days.

If you are interested in becoming a peer reviewer for *Clinical Evidence*, please complete the peer review questionnaire at www.clinicalevidence.com or contact Claire Folkes (cfolkes@bmjgroup.com).