

Qualitative research and evidence based medicine

Judith Green, Nicky Britten

Health Services
Research Unit,
London School of
Hygiene and
Tropical Medicine,
London
WC1E 7HT

Judith Green,
lecturer in sociology

Department of
General Practice,
United Medical and
Dental School of
Guy's and
St Thomas's
Hospitals, London
SE11 6SP

Nicky Britten,
*senior lecturer in
medical sociology*

Correspondence to:
Dr Green
j.green@lshtm.
ac.uk

BMJ 1998;316:1230-2

Qualitative research may seem unscientific and anecdotal to many medical scientists. However, as the critics of evidence based medicine are quick to point out, medicine itself is more than the application of scientific rules.¹ Clinical experience, based on personal observation, reflection, and judgment, is also needed to translate scientific results into treatment of individual patients.² Personal experience is often characterised as being anecdotal, ungeneralisable, and a poor basis for making scientific decisions. However, it is often a more powerful persuader than scientific publication in changing clinical practice,³⁻⁵ as illustrated by the occasional series "A patient who changed my practice" in the *BMJ*.⁶

In an attempt to widen the scope of evidence based medicine, recent workshops have included units on other subjects, including economic analysis and qualitative research.⁷ However, to do so is to move beyond the discipline of clinical epidemiology that underpins evidence based medicine. Qualitative research, in particular, addresses research questions that are different from those considered by clinical epidemiology. Qualitative research can investigate practitioners' and patients' attitudes, beliefs, and preferences, and the whole question of how evidence is turned into practice. The value of qualitative methods lies in their ability to pursue systematically the kinds of research questions that are not easily answerable by experimental methods.

We use the example of asthma treatment to illustrate how qualitative methods can broaden the scope of evidence based medicine. Although there is consensus over evidence based practice in the treatment of asthma,⁸ questions remain about general practitioners' use of clinical guidelines and patients' use of prescribed medication.⁹

Naturalism

Various qualitative methods are used in health research, but they share some basic orientations (see box).¹⁰ The first is a commitment to naturalism, or understanding health behaviour in its everyday context. Results of drug trials may inform practitioners about the optimum effects of therapeutic agents, but even pragmatic trials are not the same as everyday experience. For instance, in today's highly competitive workplace, some people report that they need to seem "healthy" at work and do not want to be seen taking medications.^{11 12} Thus, appearances may be more important to some people than symptom relief.

Summary points

Qualitative methods can help bridge the gap between scientific evidence and clinical practice

Qualitative research findings provide rigorous accounts of treatment regimens in everyday contexts

This can help us understand the barriers to using evidence based medicine, and its limitations in informing decisions about treatment

Recognising the limits of evidence based medicine does not imply a rejection of research evidence but awareness that different research questions require different kinds of research

Interpretation

A second aim of most qualitative studies is that of interpretation: investigating how patients and practitioners make sense of "objective" variables such as peak flow readings. Subjective meanings are crucial to an understanding of how treatment regimens integrate with everyday life. For instance, Adams et al found that half of the asthmatic people they interviewed did not see themselves as asthma sufferers.¹² Their "bad chests" were perceived as an acute and temporary problem, better treated with drugs to relieve the symptoms than daily prophylactic medication.

Various strategies that enable asthma sufferers to continue "normal" everyday life, despite symptoms that health professionals would see as objectively problematic, have been reported. Adams et al cite avoidance of triggers such as sexual intercourse, physical exertion, or spending time outdoors in the summer.¹² Although these adaptive strategies might seem extreme, patients interpret alternatives, such as using daily preventative medication, as accepting a stigmatising label of "asthmatic." Thus, findings about patients' possible interpretations of symptoms are essential to understanding likely medication use.

Process

The third contribution of qualitative studies to the evidence base on asthma is the assumption that social life is a process, and that interventions and the changes

Basic orientations of qualitative methods

- Naturalism—understanding treatment regimens in an everyday context
- Interpretation—what meaning do symptoms and treatment regimens have for patients and practitioners?
- Process—how might these meanings change over time?
- Interaction—how does communication between patients and practitioners impact on the meaning of medication?
- Relativism—scientific “reality” may look different from different perspectives

they bring have to be accommodated within the patient’s “biography.”¹³ Patients with asthma may have to manage medication over a whole lifetime, rather than for the limited duration of even a long drug trial, and integrate an “asthmatic” identity into their other social roles. As the study by Adams et al suggests, for many patients this will involve taking control of their medication, and limiting contact with professionals.¹² Conrad’s classic study of compliance with medication showed that patients with epilepsy often “tested” themselves to see if their epilepsy had resolved by taking themselves off medication completely or for short periods of time.¹⁴ People with asthma also have to integrate symptoms and their management with both the practicalities of everyday life and the psychological “self” over time.¹²

Interaction and relativism

Qualitative studies often take interaction as a focus of research, rather than a given factor. Katon and Kleinman viewed encounters between doctors and their patients as the bringing together of often conflicting explanatory systems about health and illness, which required negotiation to achieve good outcomes.¹⁵ The medical interview may be a meeting between very different views of reality.¹⁶ Qualitative sociologists have shown the value of a relativist approach, that focuses on these different explanatory systems.

Attitudes and adherence

One illustration of this from published reports on adherence is general attitudes to medication. Although one “reality” is that medication for asthma reduces morbidity and mortality, and can benefit users with few costs to their everyday lives, qualitative studies suggest a rather different “reality” for patients. Firstly, there is evidence that some patients have negative views about medicines, regarding these as unnatural substances that diminish the body’s own ability to fight disease and cause dependence.¹⁷ Many doctors, however, make the commonsense assumption that patients are seeking medication.¹⁸ This general finding is borne out in specific studies of patients with asthma.¹⁹ These studies show that patients worry about becoming physically and psychologically dependent on bronchodilators, and have even more deep seated worries about the long term effects of inhaled corticosteroids.¹¹ Regarding patients’ views of reality as ignorant or misguided

and attempting to persuade them of the value of a biomedical approach have limited value in increasing adherence, and the need to integrate patients’ perspectives has been recognised recently.^{15 20 21}

Assessing qualitative research

These orientations of naturalism, interpretation, process, interaction, and relativism may be shared by commonsense accounts of medical practice, such as the personal anecdote. There are, however, important differences between anecdotes (stories told for their dramatic or other qualities, without analysis or critical evaluation) and qualitative research. Rigorously conducted qualitative research is based on explicit sampling strategies, systematic analysis of data, and a commitment to examining counter explanations. Ideally, methods should be transparent, allowing the reader to assess the validity and the extent to which results might be applicable to their own clinical practice. The generalisability of qualitative research is likely to be conceptual rather than numerical.²² The studies discussed here, for instance, provide evidence for a number of conceptual issues such as the links between denial of the diagnosis of asthma and medication use or the importance of generally negative views about medication. For the practitioner, the value of these findings is that they are sensitised to issues which could usefully be explored with patients. A number of guidelines now exist to aid both journal editors and readers in assessing qualitative health research.^{10 23–25}

Conclusion

The argument that qualitative research can contribute answers to questions not easily addressed by randomised controlled trials is not new. Much has been written on the uses and value of sociology for medicine, and indeed there has been a growing acceptance of its methods in healthcare research, including its contribution to randomised controlled trials when these are appropriate.²⁶ However, in the context of the debate about evidence based medicine, it is vital to reiterate that good “evidence” goes further than the results of meta-analysis of randomised controlled trials. The



limitations of these trials, and the evidence about barriers to their utilisation by practitioners, should not lead to cynicism about the role of research evidence in health care. We need to be sure that it is the right kind of research to answer the questions posed.

Funding: No additional funding.
Conflict of interest: None.

- 1 Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet* 1995;345:840-2.
- 2 Freidson E. *Profession of medicine: a study of the sociology of applied knowledge*. New York: Dodd, Mead and Company, 1970.
- 3 Atkinson P. *Medical talk and medical work: the liturgy of the clinic*. London: Sage, 1995.
- 4 Greer A. The state of the art versus the state of the science. *Int J Technol Assess Health Care* 1988;4:5-26.
- 5 Greer A. Scientific knowledge and social consensus. *Controlled Clin Trials* 1994;15:431-6.
- 6 Burdett-Smith P. A patient who changed my practice: always check the respiratory rate. *BMJ* 1997;314:1549.
- 7 Unit for Evidence Based Practice and Policy. *Fourth UK workshop on teaching evidence-based practice study pack*. London: University College London Medical School, 1997.
- 8 British Thoracic Society and others. Guidelines for the management of asthma: a summary. *BMJ* 1993;306:776-82.
- 9 Audit Commission. *What the doctor ordered: a study of GP fundholders in England and Wales*. London: HMSO, 1996.
- 10 Mays N, Pope C, eds. *Qualitative research in health care*. London: BMJ Publishing Group, 1996.
- 11 Hewett G. *"Just a part of me": men's reflections on chronic asthma*. London: South Bank University, 1994. (Occasional papers in sociology and social policy)

- 12 Adams S, Pill R, Jones A. Medication, chronic illness and identity: the perspective of people with asthma. *Soc Sci Med* 1997;45:189-201.
- 13 Bury M. Chronic illness as biographical disruption. *Sociology of Health and Illness* 1982;4:167-82.
- 14 Conrad P. The meaning of medications: another look at compliance. *Soc Sci Med* 1985;20:29-37.
- 15 Katon W, Kleinman A. Doctor-patient negotiation and other social science strategies in patient care. In: Eisenberg L, Kleinman A, eds. *The relevance of social science to medicine*. Dordrecht: Reidel, 1981.
- 16 Mishler EG. *The discourse of medicine: dialectics of medical interviews*. Norwood, NJ: Ablex, 1984.
- 17 Britten N. Patients' ideas about medicines: a qualitative study in a general practice population. *Br J Gen Pract* 1994;44:465-8.
- 18 Hull FM, Marshall T. Sources of information about new drugs and attitudes towards drug prescribing: an international study of differences between primary care physicians. *Fam Pract* 1987;4:123-8.
- 19 Osman LM, Russell IT, Friend JAR, Legge JS, Douglas JG. Predicting patient attitudes to asthma medication. *Thorax* 1993;48:827-30.
- 20 Marinker M. Writing prescriptions is easy. *BMJ* 1997;314:747-8.
- 21 Tuckett D, Boulton M, Olson C, Williams A. *Meetings between experts: an approach to sharing ideas in medical consultations*. London: Tavistock, 1985.
- 22 Fitzpatrick R, Boulton M. Qualitative methods for assessing health care. *Quality in Health Care* 1994;3:107-113.
- 23 Dowell J, Huby G, Smith C, eds. *Scottish consensus statement on qualitative research in primary health care*. Dundee: Tayside Centre for General Practice, 1995.
- 24 Boulton M, Fitzpatrick R, Swinburn C. Qualitative research in health care. II. A structured review and evaluation of studies. *J Eval Clin Pract* 1996;2:171-9.
- 25 Blaxter M. Criteria for the evaluation of qualitative research papers. *Med Sociol News* 1996;22:68-71.
- 26 Oakley A. Who's afraid of the randomised controlled trial? Some dilemmas of the scientific method and good research practice. In: *Women's health counts*. London: Routledge, 1990.

(Accepted 21 November 1997)

North of England evidence based guidelines development project: methods of developing guidelines for efficient drug use in primary care

Martin Eccles, Nick Freemantle, James Mason

Centre for Health Services Research,
University of Newcastle upon Tyne,
Newcastle upon Tyne
NE2 4AA
Martin Eccles,
professor

Centre for Health Economics,
University of York,
York YP1 5DD
Nick Freemantle,
senior research fellow
James Mason,
research fellow

Correspondence to:
Professor Eccles
Martin.Eccles@ncl.ac.uk

BMJ 1998;316:1232-5

Practice guidelines are valid if "they lead to the health gains and costs predicted for them."¹ When implemented, valid guidelines lead to changes in clinical practice and improvements in outcomes for patients.²⁻⁵ Invalid guidelines, however, may lead to the use of ineffective interventions that waste resources, or even to harm.

Guidelines must offer recommendations for both effective and efficient care, and these have not previously been available in the United Kingdom. We have reported the development and content of guidelines for primary care in the United Kingdom based explicitly on evidence of effectiveness.⁶⁻⁹ Here, we present the methods used to develop evidence based guidelines on the use in primary care of four important groups of drugs—angiotensin converting enzyme inhibitors in patients with heart failure, choice of antidepressants, non-steroidal anti-inflammatory drugs in patients with osteoarthritis, and aspirin as an antithrombotic agent.¹⁰⁻¹³ Abridged versions of the guidelines on angiotensin converting enzyme inhibitors, aspirin, and non-steroidal anti-inflammatory drugs will be published in subsequent articles.¹⁴⁻¹⁶

Guideline development groups

Guideline development groups comprised three broad classes of members—relevant healthcare professionals (up to five general practitioners (all with an interest and

Summary points

Guideline development groups defined important clinical questions, produced search criteria, and drew up protocols for systematic review and, where appropriate, meta-analysis

Medline and Embase were searched for systematic reviews and meta-analyses, randomised trials, quality of life studies, and economic studies

Meta-analysis was used extensively by the group to answer specific clinical questions

Statements on evidence were categorised in relation to study design, reflecting their susceptibility to bias

Strength of recommendations was graded according to the category of evidence and its applicability, economic issues, values of the guideline group and society, and the groups' awareness of practical issues

Recommendations cease to apply in December 1999, by which time relevant results that may affect recommendations may be known

postgraduate training in primary care therapeutics), up to two hospital consultants, a health authority medical or pharmaceutical adviser, and a pharmacist); specialist resources (an epidemiologist (NF) and a health economist (JM)); and a specialist in guideline methodology and in leading small groups (ME). All group members were offered reimbursement of their travelling expenses and general practitioners could also claim for any expenses incurred in employing a locum.

Evidence: identification and overview

As a first step, the guideline development groups defined a set of clinical questions within the area of the guideline. This ensured that the guideline development work outside the meeting focused on issues that practitioners considered important and produced criteria for the search and the protocol for systematic review and, where appropriate, meta-analysis.

Search strategy

Searches were undertaken using Medline and, where appropriate, Embase. Using a combination of subject heading and free text terms, the search strategies located systematic reviews and meta-analyses, randomised trials, quality of life studies, and economic studies. Further details of the specific search strategies are provided in the full versions of the guidelines.¹⁰⁻¹³ Recent, high quality review articles and bibliographies and contacts with experts were used extensively. New searches were concentrated on areas where existing systematic reviews were unable to provide valid or up to date answers. The search strategy was backed up by the expert knowledge and experience of group members.

Synthesising published reports

We assessed the quality of relevant studies retrieved and their ability to provide valid answers to the questions posed. Assessment of the quality of studies considered issues of internal, external, and construct validity.¹⁷ The criteria used are shown in the box. Once individual papers had been assessed for methodological rigour and clinical importance, the information was synthesised.

Criteria for assessing quality of randomised trials

- Appropriateness of inclusion and exclusion criteria
- Concealment of allocation
- Blinding of patients
- Blinding of health professionals
- Objective or blind method of data collection
- Valid or blind method of data analysis
- Completeness and length of follow up
- Appropriateness of outcome measures
- Statistical power of results

Describing evidence

We used meta-analysis to summarise and describe the results of studies, conducting analyses to answer specific questions raised by the guideline development groups. Our primary aim was to provide valid estimates of treatment effects using approaches that provided results in a form that could best inform treatment recommendations.



MARK HUDSON

Meta-analyses combine statistically the results from similar studies and provide a weighted average of study estimates of effect. The most important criterion for combining studies is that their combination makes practical sense and, therefore, the results are interpretable. Statistical analysis procedures for meta-analysis using different outcomes are essentially analogous; all involve large sample theory and differ mainly in the details of calculations of standard errors and bias correction.¹⁸ Fixed effects models assume a common underlying effect and weight each study by the inverse of the variance. Random effects models assume a distribution of effects and incorporate this heterogeneity into the overall estimate of effect and its precision.^{19 20} Decisions on the appropriateness of fixed or random effects models were based primarily on a priori assumptions about the construct being tested in each case. Where heterogeneity between studies was identified, we also reported routinely random effects results.

Publication bias

Publication bias and missing data can undermine substantially the validity of meta-analyses.²¹ Besides using sensitive search strategies, we went to considerable lengths to obtain missing data from the trials identified. We wrote to investigators and the companies sponsoring them, and followed up non-respondents with further letters and, where appropriate, other forms of communication.

Binary outcomes

Meta-analysis of binary data, such as the number of deaths in a randomised trial, enables the results of a group of trials to be expressed in several ways (box). The pooled odds ratio is a statistically robust measure but is hard to interpret clinically; risk ratios are easier to interpret. Both are inadequate for exploring the practical implications of interventions in primary care. Risk differences are less helpful for exploring underlying effects, but are useful for describing the importance of the effects of an intervention in practice. Pooled risk differences can be adjusted for time of exposure when reviews include trials of varying lengths. This provides estimates of annual risk that can also be expressed as numbers needed to treat.²²

Meta-analysis of binary data

	Dead	Alive
Intervention group	A	B
Control	C	D

Odds ratios are defined as $\frac{A}{B} + \frac{C}{D}$

Risk ratios are defined as $\frac{A}{A+B} + \frac{C}{C+D}$

Risk differences are defined as $\frac{A}{A+B} - \frac{C}{C+D}$

Worked example

In the study of left ventricular disease treatment trial of enalapril in patients with heart failure, there were 452 deaths in 1285 patients randomised to receive enalapril and 510 deaths in 1284 patients allocated to placebo at the end of the four years' follow up.^{23,24} In a two by two table these data provide an odds ratio of 0.82, a risk ratio of 0.89, and a risk difference of -0.045 (or a 4.5% reduction in the risk of death).

Continuous outcomes

Where continuous outcomes are measured similarly in different studies, meta-analysis can be used to calculate a weighted mean difference. If measurement between studies is not undertaken using a common metric—because different instruments are used or poor reliability between those undertaking rating is likely—standardised scores based on variance within the study may be calculated for each trial. This approach, for example, enables the statistical pooling of outcomes expressed in different versions of the Hamilton depression rating scale, in which the 17 and 21 item forms are commonly used. We used the straightforward approach proposed by Hedges and Olkin,²⁵ in which the variance estimate is based upon the intervention and control group and the effect size is corrected for bias due to small sample size.

Economic analysis

The guidelines include systematic appraisals of effectiveness, compliance, safety, health service resource use, and costs of medical interventions in British general practice. The economic analysis is presented in a straightforward manner, showing the possible bounds of cost effectiveness that may result from treatment. Lower and higher estimates of cost effectiveness reflect the available evidence and the concerns of the guideline development group. Economic analyses are susceptible to bias through the methods used; we avoided making strong statements where uncertainty existed. However, the simplicity of presentation permits simple reworking with different values from the ones used by the group. This practice reflects the desire of group members for understandable and robust information upon which to base recommendations.

Presenting a review of previous economic analyses which have adopted a variety of differing perspectives, analytic techniques, and baseline data was not considered helpful. However, economic reports were reviewed to compare findings of the guideline project with representative published economic analyses and to interpret differences when these occurred.

Categorising evidence

Summarised evidence was categorised according to study design, and reflects susceptibility to bias. The box shows the categories in descending order of importance. Categories of evidence were adapted from the classification of the United States Agency for Health Care Policy and Research.²⁶ Questions were answered using the best evidence available. If, for example, a question on the effect of an intervention could be answered by category I evidence, then studies of weaker design (controlled studies without randomisation) were not reviewed. This categorisation is most appropriate to questions of causal relations. Similar taxonomies for other types of research question do not yet exist.

Categories of evidence

- Ia—Evidence from meta-analysis of randomised controlled trials
- Ib—Evidence from at least one randomised controlled trial
- IIa—Evidence from at least one controlled study without randomisation
- IIb—Evidence from at least one other type of quasi-experimental study
- III—Evidence from descriptive studies, such as comparative studies, correlation studies and case-control studies
- IV—Evidence from expert committee reports or opinions or clinical experience of respected authorities, or both

Strength of recommendation

Informal consensus methods were used to derive recommendations, and reflect the certainty with which the effectiveness and cost effectiveness of a medical intervention can be recommended. Recommendations are based upon consideration of the following: the strength of evidence, the applicability of the evidence to the population of interest, economic considerations, values of the guideline developers and society, and guideline developers' awareness of practical issues. While the process of interpreting evidence inevitably involves value judgments, we clarified the basis of these judgments as far as possible by making this process explicit. The relation between the strength of a recommendation and the category of evidence is shown in the box.

Areas without evidence

Informal consensus methods were used to develop recommendations in areas where there was no

Strength of recommendation

- A—Directly based on category I evidence
- B—Directly based on category II evidence or extrapolated recommendation from category I evidence
- C—Directly based on category III evidence or extrapolated recommendation from category I or II evidence
- D—Directly based on category IV evidence or extrapolated recommendation from category I, II or III evidence

evidence. This process sometimes identified important unanswered research questions. These are recorded at the end of the relevant section of the guideline.

Review of the guideline

External

External reviewers were chosen to reflect three groups: potential users of the guidelines, experts in the subject area, and guideline methodologists. Although the reviewers' comments influenced the style and content of the guidelines, these remained the responsibility of the development group.

Scheduled review

The recommendations of these guidelines cease to apply at the end of 1999, by which time new, relevant results that may affect recommendations are likely to be available.

We thank the following for their contribution to the functioning of the guidelines development group and the development of the practice guideline: Janette Boynton, Anne Burton, Julie Glanville, Susan Mottram.

Funding: The development of the guideline was funded by the Prescribing Research Initiative of the Department of Health. Conflict of interest: None.

- Institute of Medicine. In: Field MJ, Lohr KN, eds. *Guidelines for clinical practice. From development to use*. Washington, DC: National Academy Press, 1992.
- Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317-22.
- Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines. I. Developing scientifically valid guidelines. *Qual Health Care* 1993;2:243-8.
- Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines. II. Ensuring guidelines change medical practice. *Qual Health Care* 1994;3:552-8.
- Implementing clinical practice guidelines: can guidelines be used to improve clinical practice?* Leeds: University of Leeds, 1994. (Effective Health Care Bulletin No 8.)
- Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, et al. Developing valid guidelines: methodological and procedural issues from the North of England evidence based guideline development project. *Qual Health Care* 1996;5:44-50.
- Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, et al. North of England evidence based guideline development project: methods of guideline development. *BMJ* 1996;312:760-2.
- North of England Stable Angina Guideline Development Group. North of England evidence based guidelines development project: summary version of evidence based guideline for the primary care management of stable angina. *BMJ* 1996;312:827-32.
- North of England Asthma Guideline Development Group. North of England evidence based guidelines development project: summary version of evidence based guideline for the primary care management of asthma in adults. *BMJ* 1996;312:762-6.
- North of England Evidence Based Guideline Development Project. *Evidence based clinical practice guideline: ACE inhibitors in the primary care management of adults with symptomatic heart failure*. Newcastle upon Tyne: Centre for Health Services Research, 1997.
- North of England Evidence Based Guideline Development Project. *Evidence based clinical practice guideline: the choice of antidepressants for depression in primary care*. Newcastle upon Tyne: Centre for Health Services Research, 1997.
- North of England Evidence Based Guideline Development Project. *Evidence based clinical practice guideline: the use of aspirin for the secondary prophylaxis of vascular disease in primary care*. Newcastle upon Tyne: Centre for Health Services Research, 1997.
- North of England Evidence Based Guideline Development Project. *Evidence based clinical practice guideline: non-steroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of pain believed to be due to degenerative arthritis*. Newcastle upon Tyne: Centre for Health Services Research, 1997.
- Eccles M, Freemantle N, Mason J for the North of England ACE-inhibitor Guideline Development Group. North of England evidence based development project: guideline for angiotensin converting enzyme inhibitors in primary care management of adults with symptomatic heart failure. *BMJ* (in press).
- Eccles M, Freemantle N, Mason J for the North of England ACE-inhibitor Guideline Development Group. Evidence based guideline for the use of aspirin for the secondary prophylaxis of vascular disease in primary care. *BMJ* (in press).
- Eccles M, Freemantle N, Mason J for the North of England Non-steroidal Anti-inflammatory Drug Guideline Development Group. Evidence based guideline for the use of non-steroidal anti-inflammatory drugs versus basic analgesia in the treatment of pain believed to be due to degenerative arthritis. *BMJ* (in press).
- Cook TD, Campbell DT. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally, 1979.
- Hedges LV. Meta-analysis. *J Educ Stat* 1992;17:279-96.
- DerSimonian R, Laird N. Meta analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-88.
- Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. *Stat Med* 1991;10:1665-77.
- Freemantle N, Mason JM, Haines A, Eccles MP. CONSORT: an important step toward evidence-based health care. *Ann Intern Med* 1997;126:81-3.
- Ioannidis JPA, Cappelleri JC, Lau J, Skolnik PR, Melville B, Chalmers TC, et al. Early or deferred zidovudine therapy in HIV-infected patients without an AIDS-defining illness. *Ann Intern Med* 1995;122:856-66.
- Hedges LV, Olkin I. *Statistical methods for meta-analysis*. London: Academic Press, 1985.
- United States Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. *Acute pain management: operative or medical procedures and trauma*. Rockville, MD: Agency for Health Care Policy and Research Publications, 1992.
- SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med* 1992;327:685-91.
- SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 1991;325:293-302.

(Accepted 11 December 1997)

A rare opportunity

I saved my father's life

Fifteen years ago I was working as the only consultant surgeon in a small hospital with basic facilities in the northern part of Sri Lanka. One weekend, after my duties at the hospital, I paid a visit to my parents, who lived close to my hospital. As it was not safe to travel late at night due to the ongoing war at that time, I decided to spend the night with my parents. At midnight, I was woken up by my mother, informing me that my father had developed sudden severe pain in his inguinal hernia and started vomiting. My father had previously declined surgery for the hernia. It was not difficult for me to realise that the hernia was strangulated and needed immediate surgery. Being the only surgeon available at that time, I had no choice other than to do the surgery myself. I sent for the anaesthetist on call, and arranged the theatre for the emergency operation. In the changing room, though I have operated on many hundreds of elderly patients in that hospital, I was nervous of the outcome, with varying thoughts flashing in my mind. My father was

72, though relatively healthy. "Will there be any anaesthetic complication, bearing in mind the basic facilities available—with no electrocardiogram, monitoring, or pulse oximeter? Will there be any gangrenous bowel needing resection? Will my father survive the surgery? If not, how will I face my mother and relatives?" Quickly, I composed myself, leaving the emotions behind. Fortunately, the strangulation was relieved in time without a need for a resection of the bowel, and he made a full recovery. Everyone including the relatives praised me for my bravery to operate on my father. But who else, except myself, will know about the nervous moments I had? Since then my father has been boasting about the successful surgery by his son to all his friends. I will always feel proud and happy to be given the rare opportunity to treat and save my father, who encouraged me to study medicine.

M Vetpillai, *staff grade surgeon, Watford*

What's wrong with Bonferroni adjustments

Thomas V Perneger

Institute of Social and Preventive Medicine, University of Geneva, CH-1211 Geneva 4, Switzerland
Thomas V Perneger, *medical epidemiologist*

Correspondence to: Dr Perneger, pernegert@cmu.unige.ch

BMJ 1998;316:1236-8

When more than one statistical test is performed in analysing the data from a clinical study, some statisticians and journal editors demand that a more stringent criterion be used for "statistical significance" than the conventional $P < 0.05$.¹ Many well meaning researchers, eager for methodological rigour, comply without fully grasping what is at stake. Recently, adjustments for multiple tests (or Bonferroni adjustments) have found their way into introductory texts on medical statistics, which has increased their apparent legitimacy.²⁻³ This paper advances the view, widely held by epidemiologists, that Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference.⁴⁻⁵

Adjustment for multiple tests

Bonferroni adjustments are based on the following reasoning.¹⁻³ If a null hypothesis is true (for instance, two treatment groups in a randomised trial do not differ in terms of cure rates), a significant difference ($P < 0.05$) will be observed by chance once in 20 trials. This is the type I error, or α . When 20 independent tests are performed (for example, study groups are compared with regard to 20 unrelated variables) and the null hypothesis holds for all 20 comparisons, the chance of at least one test being significant is no longer 0.05, but 0.64. The formula for the error rate across the study is $1 - (1 - \alpha)^n$, where n is the number of tests performed. However, the Bonferroni adjustment deflates the α applied to each, so the study-wide error rate remains at 0.05. The adjusted significance level is $1 - (1 - \alpha)^{1/n}$ (in this case 0.00256), often approximated by α/n (here 0.0025). What is wrong with this statistical approach?

Problems

Irrelevant null hypothesis

The first problem is that Bonferroni adjustments are concerned with the wrong hypothesis.⁴⁻⁶ The study-wide error rate applies only to the hypothesis that the two groups are identical on all 20 variables (the universal null hypothesis). If one or more of the 20 P values is less than 0.00256, the universal null hypothesis is rejected. We can say that the two groups are not equal for all 20 variables, but we cannot say which, or even how many, variables differ. Such information is usually of no interest to the researcher, who wants to assess each variable in its own right. A clinical equivalent would be the case of a doctor who orders 20 different laboratory tests for a patient, only to be told that some are abnormal, without further detail. Thus, Bonferroni adjustments provide a correct answer to a largely irrelevant question.

Inference defies common sense

Bonferroni adjustments imply that a given comparison will be interpreted differently according to how many other tests were performed. For example, the

Summary points

Adjusting statistical significance for the number of tests that have been performed on study data—the Bonferroni method—creates more problems than it solves

The Bonferroni method is concerned with the general null hypothesis (that all null hypotheses are true simultaneously), which is rarely of interest or use to researchers

The main weakness is that the interpretation of a finding depends on the number of other tests performed

The likelihood of type II errors is also increased, so that truly important differences are deemed non-significant

Simply describing what tests of significance have been performed, and why, is generally the best way of dealing with multiple comparisons

difference in remission rates between two chemotherapeutic treatments could be interpreted as statistically significant or not depending on whether or not survival rates, quality of life scores, and complication rates were also tested. In a clinical setting, a patient's packed cell volume might be abnormally low, except if the doctor also ordered a platelet count, in which case it could be deemed normal. Surely this is absurd, at least within the current scientific paradigm. Evidence in data is what the data say—other considerations, such as how many other tests were performed, are irrelevant.

Increase in type II errors

Type I errors cannot decrease (the whole point of Bonferroni adjustments) without inflating type II errors (the probability of accepting the null hypothesis when the alternative is true).⁴ And type II errors are no less false than type I errors. In clinical practice, if a high concentration of creatine kinase were considered compatible with "no myocardial infarction" by virtue of a Bonferroni adjustment, the patient would be denied appropriate care. In research, an effective treatment may be deemed no better than placebo. Thus, contrary to what some researchers believe, Bonferroni adjustments do not guarantee a "prudent" interpretation of results.

What tests should be included?

Most proponents of the Bonferroni method would count at least all the statistical tests in a given report as a basis for adjusting P values. But how about tests that were performed, but not published, or tests published in other papers based on the same study? If several

papers are planned, should future ones be accounted for in the first publication? Should we worry about error rates related to an investigator—taking the number of tests he or she has done in their lifetime into consideration⁶—or error rates related to journals? Should confidence intervals, which are not statistical tests, but are often interpreted as such (the confidence interval includes 0, hence the groups do not differ) be counted? No statistical theory provides answers for these practical issues.

A futuristic scenario

What would happen to biomedical research if Bonferroni adjustments became routine? Cynical researchers would slice their results like salami, publishing one P value at a time to escape the wrath of the statistical reviewer. Idealists would conduct studies to examine only one association at a time—wasting time, energy, and public money. Meta-analysts would go out of business, since a pooled analysis would invalidate retrospectively all original findings by adding more tests to be adjusted for. Journals would have to create a new section entitled “P value updates,” in which P values of previously published papers would be corrected for newly published tests based on the same study. And so on . . .

Back to the Neyman-Pearson theory

These objections seem so compelling that the reader may wonder why adjustments for multiple tests were developed at all. The answer is that such adjustments are correct in the original framework of statistical test theory, proposed by Neyman and Pearson in the 1920s.⁷ This theory was intended to aid decisions in repetitive situations. Imagine that your factory produces light bulbs in lots of 1000, and that testing each bulb before shipment would be impractical. You can decide to test only a sample in each lot, and to reject

(literally) any lots in which more than a predefined number (x) of bulbs in the sample are defective. Of course, your decision might be wrong for any particular lot, but the Neyman-Pearson theory provides a decision rule (the number x), so that over many trials your error rates (type I and type II) will be minimised. Now, if for some reason you took 20 samples out of a given lot instead of one, and decided that you would reject the lot if the number of defective bulbs exceeded x in only one sample, you would be much too likely to reject a good lot in error, and a Bonferroni adjustment would restore the original optimal error rates.

The catch is that Neyman and Pearson developed their statistical tests to aid decision making, not to assess evidence in data. The latter practice may be objected to for several reasons (this topic would deserve a discussion of its own), and alternative approaches to statistical inference, such as estimation procedures, use of likelihood ratios, and Bayesian methods, have been proposed.^{8–11} Bonferroni adjustments follow the original logic of statistical tests as supports of repeated decisions, but they are of little help in determining what the data say in one particular study.

Should Bonferroni adjustments ever be used?

Statistical adjustment for multiple tests make sense in a few situations. Firstly, the universal null hypothesis is occasionally of interest. For instance, to verify that a disease is not associated with an HLA phenotype, we may compare available HLA antigens (perhaps 40) in a group of cases and controls. If no association existed, at least one test would be significant with a probability of 0.87, and Bonferroni adjustments would protect against making excessive claims. A clinical equivalent is the case of a healthy person undergoing several laboratory tests as part of a general health check. Secondly, adjustments are appropriate when the same test is repeated in many subsamples, such as when stratified analyses (by age group, sex, income status, etc) are conducted without an a priori hypothesis that the primary association should differ between these subgroups. Note that this is the scenario, reminiscent of repeated sampling of the same lot, that Tukey and Bland and Altman use in their justifications of multiple test adjustments.^{1–3} Sequential testing of trial results also falls in this category. A final situation in which Bonferroni adjustments may be acceptable is when searching for significant associations without pre-established hypotheses.

The best approach

However, even in these situations, simply describing what was done and why, and discussing the possible interpretations of each result, should enable the reader to reach a reasonable conclusion without the help of Bonferroni adjustments.^{5–12} There is an important difference between what the data say and what the researcher (or the reader) believes to be true.⁸ The latter depends not only on the data at hand but also on considerations such as whether a finding is biologically plausible or whether the significant test was a serendipitous finding in a fishing expedition. The



integration of prior beliefs with evidence is best achieved by Bayesian methods, not by Bonferroni adjustments. In summary, Bonferroni adjustments have, at best, limited applications in biomedical research, and should not be used when assessing evidence about specific hypotheses.

I thank Dr Richard M Royall, Department of Biostatistics, Johns Hopkins University, for helpful comments on the manuscript.

Funding: Swiss National Science Foundation (PROSPER 3233-32609.91).

Conflict of interest: None.

- 1 Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
- 2 Greenhalgh T. Statistics for the non-statistician. I. Different types of data need different statistical tests. *BMJ* 1997;315:364-6.
- 3 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;310:170.

- 4 Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990;1:43-6.
- 5 Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *Am J Epidemiol* 1995;142:904-8.
- 6 Thomas DC, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong RG. The problem of multiple inference in studies designed to generate hypotheses. *Am J Epidemiol* 1985;122:1080-95.
- 7 Neyman J, Pearson ES. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 1928;20A:175-240, 263-97.
- 8 Royall RM. *Statistical inference: a likelihood paradigm*. New York: Chapman and Hall, 1997.
- 9 Berkson J. Tests of significance considered as evidence. *J Am Stat Assoc* 1942;37:325-35.
- 10 Goodman SN, Royall R. Evidence and scientific research. *Am J Public Health* 1988;78:1568-74.
- 11 Oakes M. *Statistical inference*. Boston: Epidemiologic Resources, 1990.
- 12 Jones DR, Rushton L. Simultaneous inference in epidemiologic studies. *Int J Epidemiol* 1982;11:276-82.

(Accepted 16 January 1998)

Coping with loss

The doctor's losses: ideals versus realities

Glin Bennet

This is the sixth in a series of 10 articles dealing with the different types of loss that doctors will meet in their practice

After five years of study, newly qualified doctors may find it hard to realise that much of their future development will involve loss. They will go on gathering information and acquiring skills, but if they are to retain their enthusiasm and to mature as people, they will be learning to live with various losses.

Tiredness

New doctors should enjoy the initial enthusiasm, the ideals and the sense of omnipotence and invulnerability, the buoyant feeling of being able to contribute to the general good, because it may not last for long. Very likely a few months of broken nights will blur the ideals and push the ambitions into the distance. The immediate objective becomes to get through the job.

The grinding tiredness teaches them a lot: about their limitations, that sleep matters, and that it is difficult to be a good doctor when their eyes will not stay open. They become impatient over explanations, and tiredness comes up like a barrier so that they can no longer reach out to anxious and grieving patients.

They are learning that they cannot meet the ideals they set for themselves or the expectations of others. But tiredness is cured by a good sleep and enthusiasm is restored by a relaxing weekend. They can be admired for the long hours they work. They work harder than other people, they work amid the basic crises of living, they know about suffering, they see that people get better through their individual efforts, though they are not successful all the time. The death of a patient is a loss that reminds doctors of their limitations and the limitations of medical science, in which they had been taught to have so much faith. The first time it happens, the doctor is sad, shocked, perhaps angry that the patient could have done that to them.

Summary points

Reality often disappoints the expectations of young doctors, who become tired and disillusioned with themselves and with the health care system

A plateau in middle life is often associated with loss of further opportunities, and high achievers may interpret this as failure

To enjoy medicine we must achieve a balance between meeting the needs of our patients and maintaining our own resources of strength, energy, and commitment

Doctors who can acknowledge their own fallibility, accept their own wounds, and accept help from colleagues or others may emerge warmer and more humane

Loss of unreality

Most doctors have relatively simple lives in these early years, so it is possible, if they want, to give all their waking hours to the work in hand. Then there comes a time when the work is not sufficiently sustaining on its own—at least it ceases to be for most people, especially when the needs of others have to be considered. Now the people with the idealism and enthusiasm are confronted with a fresh reality, and much of a doctor's subsequent life and career will depend on how this matter is addressed.

This is a further lesson in the loss of omnipotence, but in no way is it the beginning of a decline. It is a time for redirecting energy. Doctors who accomplish this and can control the circumstances of their work can have a satisfying life, because medicine offers such abundant opportunities.

44 Wellington Park,
Bristol BS8 2UW
Glin Bennet,
formerly consultant
psychiatrist, United
Bristol Healthcare
NHS Trust

Glin.Bennet@
quadbtinternet.com

Series editors:
Colin Murray
Parkes and Andrew
Markus

BMJ 1998;316:1238-40

If, as doctors, we are to retain our enthusiasm and to mature as people, we must learn to live with loss

Many doctors make choices that put them in the front line, where they are directly exposed to needy members of the public. Here, the external pressures may seem always to be about to overwhelm them. The ideals are abandoned, and the redirection cannot be achieved. The giving out exceeds the individual's restorative powers. If as established doctors we find ourselves perpetually rushing to catch up with the demand as we see it, if our families tell us that we are irritable and our friends that we look tired all the time, then we are letting our resources become depleted: the signs of burnout will appear.

There is a phrase: "You have to be on fire before you can burn out." The idealism is gradually replaced by a mild cynicism, patients are perceived as inconsiderate and ungrateful, the telephone becomes an enemy. The process is familiar and, if unchecked, leads on to an apathy, in which minimum energy goes into the clinical work, although it may be redirected into administrative and extracurricular activities.

Men and women have different experiences

The traditional male approach has been to disregard these issues, at the outset of his career or later, and carry on up the professional ladder, regardless of personal considerations. If he becomes powerful enough he can sustain this style of life and put off his maturing, possibly forever. Sadly, these are qualities that make for professional success, and such people can be found among the more influential teachers of medical students, who are thereby exposed to the model of brashness and emotional immaturity.

Women doctors often experience loss before they qualify. They are affected more than most men by the brutalising aspect of medical education, which diminishes the empathic part of clinical work. They delay childbearing to their late 20s or beyond, and then they have to work that much harder to achieve the same professional goals as men.

Loss of meaning, loss of spirit

In the professional man's progress there are critical points, or hurdles, that can be cleared successively and that set him fair for the next stage. The loss of omnipotence is a necessary first stage. Another critical time for men comes in their late 30s, when they usually have their consultant post or their partnership in general practice. They are probably married with children, have a house, a good car, and a boat as well perhaps. The challenge here is that there are no more explicit challenges, such as qualifications and jobs; they have all been attained. Thus, there is a loss of these defining events and the excitement about all that is one day going to be. The doctor wakes one morning and says: "My life is now. This is what I am." This is the plateau of middle life.

High achievers imagine, albeit subconsciously, that the only way on from the plateau is downhill, and that once they have reached a particular professional peak there can only be loss of status, loss of role, diminishing health, and so on. The losses are real, but like all losses

they are points of transition, which can be seized creatively. There are gains from this new state: it is no longer necessary to meet all the expectations of others, to keep on achieving or publishing; it is no longer necessary to be wise and in control. There is more time. Life can broaden out at this point, provided that the losses are understood and seen for what they really are.

Women suffer loss in the process of becoming doctors, but there are great benefits later on, and unfulfilled male doctors could learn from their female colleagues. A woman doctor has professional work that she enjoys and often she has a family; thus she is not looking solely to her job to give her a life that is meaningful. Male doctors can do the same, only so often they avoid their families.

Doctors going wrong

When their life can no longer be defined by achievement, doctors often fail to cope. The work has lost its meaning because there is simply too much of it or because the doctor has not adapted to changing circumstances. The issues of omnipotence and invulnerability appear again, but in a more subtle form: it is no longer a matter of one not being able to accomplish everything, but rather accepting that one is merely a vulnerable human being. This comes hard to powerful doctors, and they may try to avoid the transition in various ways: by blotting it out (through alcohol); by reassuring themselves of their potency (with a new relationship); by channelling their energies in new directions (by getting on the train to London, where the important committees meet); by developing symptoms and becoming depressed.

All losses are points of transition, which can be seized creatively

Most problems that doctors experience in middle life are essentially problems of meaning, and their lives have lost their meaning because they have not been able to make the fundamental transitions and to value themselves simply for being who they are.



ADRIANA DORSETT

Confidential advice and support

There are currently many agencies for helping doctors. One example is the BMA's Stress Counselling Service (tel 0645 200169). It offers confidential telephone counselling 24 hours a day, 365 days a year, on a wide range of personal, emotional, and work related problems in areas such as:

- Debt and other financial concerns
- Workplace problems
- Difficulties with marital and personal relationships
- Alcohol and drug misuse
- Loss of confidence
- Stress and anxiety
- Bereavement.

Helpline numbers of two other agencies are 07071 223372 and 0121 558 0278.

losses are denied by them, so they are unable to make the essential transitions. It is as if they can accept themselves only as immaculate and all-competent professionals, and any blemish on that image is seen as a failure.

The losses are losses of illusions; the gains are gains in reality

What is to be done?

It is good that schemes have been organised whereby doctors in difficulties can get confidential help at a distance from their places of work, but there is a measure of failure here. Doctors become involved in such schemes because they have not been able to share their deeper feelings with their immediate colleagues, or because problems have been neglected until the situation is out of hand and the doctor is ceasing to function competently.

These doctors are stranded. They have of course lost their omnipotence and invulnerability, and they are no longer climbing the ladder to success, but these

In fact the blemishes have the potential to be a great advantage, and these doctors would do well to attend to the ancient idea that only the wounded physician heals. In myth it is presented in a literal form, but in ordinary life the "wounded" state refers merely to the acceptance of one's imperfections.

When doctors can accept their blemishes and vulnerabilities and their inability to achieve everything, they are free to make warm and ordinary relationships with their patients, family, and friends. They are free to look at the quality of their work and to make changes where these are needed. The losses are losses of illusions; the gains are gains in reality, and the quality of work and the quality of life can improve beyond recognition.

Further reading

Bennet G. *The wound and the doctor: healing, technology and power in modern medicine*. London: Secker and Warburg, 1987.

The articles in this series are adapted from *Coping with Loss*, edited by Colin Murray Parkes and Andrew Markus, which will be published in July.

A memorable patient
Old remedy effective

On a Shetland island, the home of hardy people, it was surprising to find that a runny nose and a cough meant that a visit to the doctor was required. And this was 40 years ago when life was less comfortable than today, when television was hardly known, and when people visited each other of an evening to talk and drink "a peerie glass o' the hame brew."

I was covering the period between the departure of the island's only doctor and the arrival of his successor. It was a wonderful six weeks among wonderful people. Their attitude to the common cold, though, was a bore. Mr X came to the surgery requesting a bottle of medicine for his cough. Explanation of the self limiting nature of his illness was treated with contempt. This being a dispensing practice I handed him some proprietary cough mixture from the big cupboard. No good. In two days he was back, seeking something more effective. My advice to hang on was again abruptly treated as nonsense and he went away with something of a different colour. Again, no good. It was too soon to consider investigating his cough, and as he was as determined as ever to take medicine I turned yet again to the cornucopia in the wall. There were now no more cough mixtures. There were, however, many big stock bottles of coloured glass containing ancient remedies with Latin names, and wedged among them, an out of date *National Formulary*. I sent Mr X home, promising to see if I could concoct something for him.

Mist Morph et Ipecac, with another now forgotten ingredient, was the result, the whole being dignified by the suffix "BPC." I made this up, using strange little measuring glasses graduated in

outdated units of measurement. I poured it into a bottle, stuck a cork in it, and labelled it with the instructions. It is with pride that I recall that I then took a dose of this concoction, believing that if I survived it he would. Mr X received this medicine and I heard no more till he returned four weeks later to have a loose tooth removed. This gave much less trouble than his cough but to the latter there was to be a sequel.

I took a phone call from an isolated cottage at the south end of the island. "Hello," a woman said, "I don't want you to come down but I've got a terrible cough and I just wondered if you would give the Postie a bottle of that medicine you made for Mr X to bring down to me. He says it's the same one that old Dr S used to make 30 years ago."

It is said that the only thing that distinguishes man from the lower animals is man's desire to take medicine. Especially it seems old medicines, and even in Shetland.

Bill Yule, *retired general practitioner, Forfar*

We welcome articles up to 600 words on topics such as *A memorable patient*, *A paper that changed my practice*, *My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.