

Adverse events in health care: issues in measurement

Kieran Walshe

Introduction

Adverse events—“instances which indicate or may indicate that a patient has received poor quality care”¹—are used widely in healthcare quality measurement and improvement activities. Many commonly employed quality improvement mechanisms, such as incident reporting, occurrence screening, significant event auditing, processes for dealing with complaints, and (in the UK) the national confidential enquiries into various areas of clinical care are essentially focused on such adverse events. Even traditional medical quality improvement mechanisms such as mortality and morbidity conferences or death and complications meetings are predicated on the idea that by identifying and examining adverse events, we can learn lessons and change practice in ways that will make such events less likely in future and hence improve the quality of health care.

The principle that studying adverse events can produce information which leads to quality improvements is far from new and has been much used outside of health care.^{2,3} It has an intuitive power—after all, we all learn much as individuals from our own mistakes, and it seems reasonable to hypothesise that organisations can also learn a great deal from their errors. However, it is easy to overlook the complexities of measurement involved in defining, classifying, identifying, describing, and analysing such adverse events.⁴ Like any other measurement tools, those used with adverse events need to be tested to ensure that they work. This article presents an analysis of the issues involved in defining adverse events, the sources of data which can be used to identify such events, and the validity and reliability of measures of quality based on adverse events in health care.

The idea that it would be useful or important to study the incidence, circumstances, or causes of adverse events in health care arises from various different but related schools of thought. For example, researchers concerned with the level and impact of iatrogenic disease,^{5,6} those interested in measuring and improving quality,^{7,8} others investigating medical malpractice, negligence, and litigation,^{9,10} and some interested in the human and organisational psychology involved¹¹ have all developed approaches to studying adverse events in health care. Researchers have examined differ-

ent aspects of the epidemiology of adverse events—their consequences for patients, the costs for healthcare organisations, the perceptions of clinicians and others involved in these events, the causes and factors which contribute to their occurrence, their preventability, their use in performance measurement, and so on. Some common themes can be identified. Researchers concur that adverse events are important and worthy of study and investigation because they are quite prevalent (and more prevalent than might be expected), have important impacts on healthcare organisations and patients, and are often apparently preventable. Researchers also seem to agree that the study of adverse events should look beyond the performance of the individual clinician, and recognise the importance of the wider process of care and the organisational context in which it takes place.

Defining adverse events in health care

The starting point for any measure based on adverse events must be a definition of what constitutes such an event. Several different researchers have developed definitions for the term *adverse event*, and table 1 lists some of the principal ones.

Reviewing these definitions, it is clear that they largely agree that an adverse event is a happening, incident, or set of circumstances which exhibits three key characteristics to some degree:

- *Negativity*: it must be an event which is, by its very nature, undesirable, untoward, or detrimental to the healthcare process or to the patient. This is a theme which is common to all definitions
- *Patient involvement/impact*: it must in some way involve or have some negative impact or potential impact on a patient or patients. The wider definitions of adverse events include occurrences in which there is no actual effect on any patient, though there is the potential for harm. More restrictive definitions often only include events where the patient has suffered some definable and identifiable ill effect from the event
- *Causation*: there must be some indication that the event is a result of some part of the healthcare process (either through commission or omission), rather than a result of events outside the healthcare process, such

Correspondence to:
Dr K Walshe, senior research fellow, Health Services Management Centre, University of Birmingham, Park House, 40 Edgbaston Park Road, Birmingham B15 2RT, UK

Table 1 Some definitions of adverse events drawn from the literature

| | |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| McLamb and Huntley (1967) ¹² Mills (1978) ⁹ | <p>“Any response to medical care in the hospital that is unintended, undesirable and harmful to the patient.”</p> <p>“A potentially compensable event is a disability caused by health care management:</p> <ul style="list-style-type: none"> ● Disability—is a temporary or permanent impairment of physical or mental function (including disfigurement) or economic loss ● Causation—is established when the disability is more probably than not attributable to health care management ● Healthcare management—includes both actions and inactions of any healthcare provider or attendant.” |
| Craddick and Bader (1983, p23) ¹³ | “Adverse patient occurrences (APOs) ... refer to untoward patient events which, under optimal conditions, are not a natural consequence of the patient’s disease or treatment. The common thread of all APOs is that they are events which health professionals agree are not desirable outcomes of medical management.” |
| Harvard Medical Practice Study (1990) ¹⁰ | “An unintended injury caused by medical management rather than by the disease process. The injury is sufficiently serious to lead to prolongation of hospitalisation or temporary or permanent impairment or disability in the patient.” |
| Wilson RM <i>et al</i> (1995) ¹⁴ | “An unintended injury or complication which results in disability, death or prolonged hospital stay and is caused by health care management.” |
| Walsh (1998) ¹⁵ | “An untoward or undesirable occurrence in the healthcare process which has or potentially has some negative impact on a patient or patients and results or may result from some part of the healthcare process.” |

as the patient’s own actions or the natural progression of the disease. Again, definitions vary, with some accepting events as adverse events with little or no evidence of causation, while others insist on strong and direct evidence of causation.

The definitions listed in table 1, however, are of limited value in actual measurement because they do not define the circumstances or events which constitute an adverse event in sufficiently concrete terms to allow such events to be identified reliably. To this end, most measures of adverse events make use of a series of statements or criteria which operationalise the definition by describing a series of circumstances or instances which are seen as adverse events. Table 2 shows an example of such a list. Behind each example listed in the table might lie, in turn, a further description or definition setting out the details of what constitutes such an adverse event. Some use such a list of further definitions largely as a prompt list, and rely on the professional skills and experience of the person using the measure to decide whether a particular instance is or is not an adverse event, while others use a second stage of review by a senior clinician for this purpose. Some have endeavoured, however, to define the nature of each type of adverse event in specific terms to allow the measure to be used reliably by many different users without, necessarily, a high level of clinical expertise.

When adverse events are identified using a measure or tool such as that in table 2, they are then often analysed or categorised further in various ways. Their effect on the patient involved may be rated, in terms of their severity and temporal persistence, and their effect on the organisation and the costs of health care may also be considered. The cause of the event may be explored in an attempt to distinguish between those which arise from the healthcare system and those which may result from the underlying disease process or from other causes, and in order to attribute events to particular parts of the healthcare organisation. The avoidability of events or the acceptability of the standard of care provided may also be rated to make an explicit professional assessment of the quality of care. Sometimes, an assessment of the existence of negligence (which is a medicolegal rather than a clinical judgment) may be made.

The classification of adverse events in these ways is almost always done through some form of professional review. The rigour with which those reviews are undertaken varies—from those which are simply based on a single professional’s personal and implicit assessment of the circumstances, to those which use multiple professional assessments, made with explicit criteria and definitions of the concepts involved. Investigations of the reliability and validity of this review process have indicated

Table 2 Generic screening criteria used to identify adverse events¹³

| | |
|----|----------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Admission for complications or adverse results of outpatient management |
| 2 | Admission for complications or incomplete management of problems on previous admission to hospital |
| 3 | Operative consent incomplete, missing, or otherwise incorrect |
| 4 | Unplanned removal, injury, or repair of organ or structure during surgery, invasive procedure, or vaginal delivery |
| 5 | Unplanned return to operating room on this admission |
| 6 | Invasive procedure with tissue removed where pathology report does not match preoperative diagnosis, or non-diagnostic or no tissue removed |
| 7 | Transfusion required for bleeding/anaemia/other iatrogenic reason, not clinically indicated, or resulting in reaction |
| 8 | Nosocomial (hospital acquired) infection |
| 9 | Antibiotic or drug use problems |
| 10 | Cardiac or respiratory arrest, or low Apgar score |
| 11 | Transfer from general care unit to special care unit |
| 12 | Other patient complications |
| 13 | Hospital incurred patient incident, such as fall, intravenous problem, medication error, or skin problem |
| 14 | Abnormal laboratory, radiograph, or other test results not addressed by physicians |
| 15 | Neurological deficit present at discharge which was not present on admission |
| 16 | Transfer to another acute care facility |
| 17 | Death |
| 18 | Subsequent visit to emergency department or outpatients department for complication or adverse results related to this admission to hospital |
| 19 | Length of stay above certain centile or allotted days |
| 20 | Medical record review |
| 21 | Nursing record review |
| 22 | Departmental or other problems |
| 23 | Patient or family dissatisfaction |

that consistent intra-rater and inter-rater reliability are elusive, and that the achievement of reliable judgments may demand more multiple ratings than are practically feasible.¹⁶⁻¹⁸ Such professional reviews can also be biased by the knowledge of case outcomes.¹⁹

Sources of data on adverse events

One approach to identifying adverse events in health care is to monitor or screen patients' clinical records either during or after the process of care. Information is abstracted from the clinical records by staff who use the records to decide whether or not adverse events have occurred and to document and classify those events. There are two important weaknesses in this process. Firstly, the clinical records may be deficient, and as a result adverse events might be missed. Indeed, because the more deficient the medical records are the harder it will be to identify adverse events, the paradoxical situation could occur in which good, comprehensive records produce a higher adverse event score (and so an indication of a lower quality of care) than sketchy, incomplete records. Secondly, the clinical records are always a summary of events in the patient's care and treatment rather than a record of every action and incident. Some adverse events might concern circumstances which are not routinely recorded in the clinical record, and so reliance on the clinical record as the sole source of information might produce a spuriously low indication of their incidence.

Another source of information on adverse events is the self reporting of incidents by clinical professionals. Indeed, most healthcare organisations have at least some reporting mechanisms for a range of adverse events such as medication errors and patient accidents.^{20 21} If the reliability of the clinical records is a concern, however, the reliability of reporting mechanisms which rely on many different professionals to report adverse events, all of whom may have different personal definitions of what constitutes an adverse event and different degrees of commitment to the self reporting mechanism, must be even more in doubt. Some researchers have reported that adverse event reporting misses many adverse events which records screening would identify,^{13 22} though others have found the two methods to be equally productive.²³

With the increasing availability of information technology in hospitals, some researchers have used available data held within computer systems to identify fairly limited groups of adverse events,^{24 25} but have found that unless wholly computerised clinical records are available, most adverse events cannot be identified in this way.

When adverse events are used in quality measurement, some applications simply identify individual events or case series of similar or related events which are explored using qualitative methods,⁸ often based on the ideas of the critical incident technique.²⁶ Many make use of some kind of quantitative analysis, however, in which adverse events are counted, aggregated, converted into rates, standardised for various

factors, etc. In both cases, it is important to consider the validity and reliability of the definitions and measures being used.

Validity of adverse event measures of the quality of health care

FACE AND CONTENT VALIDITY

Face validity is a measure of whether an instrument seems reasonable, and produces reasonable data, from the viewpoint of its users. Content validity is a measure of whether the items within an instrument adequately reflect the conceptual definition of its scope.²⁷ Few studies of the face and content validity of adverse event measures exist, but those that do (table 3) suggest these measures are valid. One survey of 150 doctors in public health and clinical medicine in the UK found broad support for the validity of a generic adverse event measure, although participants suggested many improvements to the detail of the measure's definition.¹⁵ A parallel interview study found that although the principle of using adverse events in measuring quality was supported, clinicians had some concerns about the practice leading to an undue focus on such events unless other dimensions of quality were also measured. A small US study used a panel of three doctors to rate the "adversity" of each element of the adverse patient occurrences inventory,²⁸ and reported that "all of the weights obtained were negative, and the physicians generally agreed with one another in their evaluations".

CRITERION RELATED VALIDITY

Criterion related validity is a measure of the relation between measurements made using an instrument and an external variable (the criterion, sometimes called the gold standard) with which it is expected to correlate.²⁷ In assessing the criterion related validity of an adverse event measure of healthcare quality, the most important and difficult issue is the selection of an appropriate criterion. The few researchers who have studied the criterion related validity of adverse event measures have mostly used some form of implicit professional assessment of the quality of care as their criterion. Although this is obviously simpler to do than identifying a separate explicit measure of the quality of care as the criterion, the acknowledged low validity and reliability of such implicit professional judgments^{16 17 29} present some difficulties.

Studies of the criterion related validity of adverse event measures generally support their validity, with some provisos (table 3). They suggest that adverse event measures may suffer from high false positive rates—identifying many cases as having adverse events when in fact they contain no real quality problem, and that different individual adverse event definitions may have quite different validity characteristics. This implies that the validity of adverse event measures may be crucially dependent on the mixture of adverse events they contain, and that attempts to develop or test adverse event measures should pay attention to the individual items included within the instrument. They also suggest that adverse

Table 3 Summary of research into the validity of adverse event measures of healthcare quality

| Study | Dimension of validity | Methods | Results and conclusions |
|----------------------------------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Panniers and Newlander (1986) ³⁰ | Construct | Applied adverse patient occurrence inventory to sample of 426 patients in DRGs 121 and 122 (myocardial infarction) from one hospital, and collected details of length of stay and treatment costs | Significant correlation between adverse patient occurrence inventory score and both length of stay ($p < 0.001$) and treatment costs ($p < 0.001$) |
| Panniers (1987) ³¹ | Construct | Applied adverse patient occurrence inventory and AS-SCORE to sample of 354 patients in DRGs 121 and 122 (myocardial infarction) from one hospital, and also collected details of length of stay and treatment costs. | Significant correlation between adverse patient occurrence inventory score and AS-SCORE ($p < 0.001$) |
| Schumacher <i>et al</i> (1987) ³² | Construct | Applied adverse patient occurrence inventory and severity of illness index to 752 patients from 7 hospitals. | Significant correlation between APO inventory score and severity of illness index ($p < 0.0001$) |
| Barnes and Moynihan (1988) ³³ | Criterion related | Used 3252 records from SuperPRO project, drawn from PROs all over USA. Applied PRO generic screens for primary review; used single physician assessment for secondary review | Primary screening had sensitivity of 48.5%, specificity of 72.8%. Missed adverse events tended to be less clinically significant ones. Some adverse event definitions were more valid than others |
| Richards <i>et al</i> (1988) ²⁸ | Face/content | Used panel of 3 doctors to rate "adversity" of each element of adverse patient occurrence inventory | Results not specifically reported. All ratings obtained were negative and "generally agreed" |
| Richards <i>et al</i> (1988) ²⁸ | Criterion related | Applied adverse patient occurrence inventory, severity of illness index, and condition specific quality measures to 25 000 patients from 45 hospitals, and also collected administrative/demographic and clinical details of cases | Adverse patient occurrence inventory score correlated with criterion based quality measures for more serious conditions |
| Richards <i>et al</i> (1988) ²⁸ | Construct | Same data set | Adverse patient occurrence inventory score correlated with length of stay and with severity of illness index. Some variations in adverse patient occurrence inventory score observed across different types of hospitals. |
| Brennan <i>et al</i> (1989) ³⁴ | Criterion related | Applied own adverse event measure twice each to 360 cases selected to contain a high proportion of cases with adverse events; undertook a separate single physician assessment of cases | For general population of cases, estimated measure had sensitivity of 21% and specificity of 99.5% |
| Institute of Medicine (1990) ³⁵ | Criterion related | Used 6 309 839 cases reviewed by PROs up to June 1989 to compare percentages of cases with an adverse event identified at primary screening and percentage of those adverse events subsequently confirmed as quality problems | Between 7.5% and 71.3% of adverse events were confirmed as quality problems—rate varied by type of adverse event |
| Brennan <i>et al</i> (1991) ³⁶ | Construct | Used 31 000 patient records drawn from 51 hospitals to compare adverse event rates and negligent adverse event rates across hospitals | Adverse event rate and negligent adverse event rate varied substantially across hospitals. Primary teaching hospitals had higher adverse event rates; rural hospitals had lower adverse event rates. Proportion of adverse events due to negligence lower in teaching and proprietary hospitals; higher in non-profit and governmental hospitals |
| Bates <i>et al</i> (1995) ³⁷ | Criterion related | Applied adverse event measure to 3137 admissions; undertook a separate clinician review of those screened positive to assess presence, seriousness and preventability of adverse event | Sensitivity of screening criteria ranged from 68% to 4%; specificity ranged from 56% to 99%. No criteria had both high sensitivity and high specificity |
| Walshe (1998) ¹⁵ | Face/content | Surveyed 150 doctors in public health and clinical medicine in UK on validity of a general adverse event measure. Interviewed separate sample of 6 doctors | Broad support for validity, though many suggestions for improvements to the detail of the measure. No significant differences between public health and clinical doctors' views. Interview study showed support for adverse event measure but some concerns about need for other measures of quality to be used too. |
| Walshe (1998) ¹⁵ | Construct | Used data on 12 676 admissions in 8 specialties all screened for adverse events using generic and specialty specific measures | Adverse event rates varied significantly among specialties, were associated with length of stay, mode of admission, and death in hospital |

PRO = peer review organisation; DRG = diagnosis related group

event measures may provide a valid measure of quality only for those patients whose illness is relatively severe.

CONSTRUCT VALIDITY

Construct validity is a measure of how well an instrument supports or conforms with theories or constructs.²⁷ Exploring the construct validity of adverse event measures of quality is difficult because there are few established theories and constructs about the distribution and effects of adverse events for researchers to test. However, as table 3 shows, researchers have shown that adverse event rates are associated with increases in length of stay and resource usage, that different sorts of adverse events are associated with different levels of increase in length of stay, that rates of adverse events vary between specialties, that patients who are more severely ill on admission, patients who are emergency admissions, and patients who die in hospital have more adverse events, and that rates of adverse events vary between hospitals and are higher in some types of hospitals than in others. Overall, there is reasonable evidence for the

construct validity of adverse event measures of quality.

Reliability of adverse event measures of the quality of health care

Inter-rater reliability measures whether, when the same test is applied to the same respondent or subject by different raters, the same results are produced.²⁷ Several researchers have examined the inter-rater reliability of adverse event measures, generally by arranging for multiple reviews of patients' case records by different screening staff and then comparing the results of screening. Their results (summarised in table 4) are mixed, but most studies indicate that the reliability of adverse event measures is at best moderate to good, and that the reliability of measurement may be highly dependent on the quality of rater training and ongoing monitoring as well as the construction of the measure.

Intra-rater reliability measures whether, when the same test is applied to the same respondent or subject by the same rater on two different occasions, the same results are

Table 4 Summary of research into the reliability of adverse event measures of healthcare quality

| Study | Dimension of reliability | Methods | Results and conclusions |
|-----------------------------------------------------|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Panniers and Newlander (1986) ³⁰ | Inter-rater reliability | Used 2 raters to apply modified form of adverse patient occurrence inventory to sample of 200 cases from 426 patients with myocardial infarctions | Raw agreement of 99–100% for 10 items of adverse patient occurrence inventory; other 5 items ranged 72–96% (κ 0.29 to 0.83). Concluded adverse patient occurrence inventory was reliable |
| Schumacher <i>et al</i> (1987) ³² | Inter-rater reliability | Used 7 raters to apply adverse patient occurrence inventory to 752 cases (each being reviewed 2 or 3 times) drawn from 7 hospitals | Pearson correlation coefficients cited, measuring association between raters. Mean correlation for adverse patient occurrence score was 0.33 (ranged from -0.05 to 0.58). Concluded adverse patient occurrence inventory insufficiently reliable |
| Richards <i>et al</i> (1988) ²⁸ | Inter-rater reliability | Used multiple raters to apply adverse patient occurrence inventory to 516 cases drawn from 5 hospitals, each reviewed by 2 raters | K statistics for adverse patient occurrence numerator items had mean of 0.33 (ranged -0.18 to 0.73); for adverse patient occurrence denominator items mean was 0.50 (ranged 0.28 to 0.83). For adverse patient occurrence score, found within-patient variability much less than overall variability. ANOVA showed differences between raters responsible for 2% of adverse patient occurrence score variability. Concluded adverse patient occurrence inventory “at best moderately reliable” |
| Harvard Medical Practice Study (1990) ¹⁰ | Inter-rater reliability | Used multiple raters to apply own adverse event measure to 282 cases (random 1% sample of total study), each reviewed by 2 raters | Raw agreement on presence/absence of adverse event in each case of 93.6%, κ of 0.85. Concluded measure was sufficiently reliable for use in study |
| Walshe (1998) ¹⁵ | Inter-rater reliability | Used multiple raters to apply adverse event measure to 374 admissions across three specialties, each reviewed by 2 raters | Overall κ statistics of 0.46, 0.63, and 0.65 in three specialties, suggesting “moderate to good reliability” but much dependent on rater training |
| Walshe (1998) ¹⁵ | Intra-rater reliability | Used a single rater to apply adverse event measure to 110 admissions in obstetrics, and then rescreened same records 4 months later | Overall κ statistic of 0.56 suggesting moderate reliability. Significantly more adverse events found on second screening when rater aware study being undertaken |
| Walshe (1998) ¹⁵ | Inter-rater reliability | Observational study of 6095 admissions in 8 specialties screened by four different raters for adverse events | Significant differences in rates of adverse events detected by different raters found in 6 specialties |

produced. One study has examined intra-rater reliability and has found moderate reliability.

Conclusions

Adverse events are clearly important to healthcare organisations, not only because of their impact on patients but also because they can provide an insight into the quality of health care and an opportunity for improvement. Adverse events can, as individual instances of care, provide an information-rich and compelling case for action and improvement, and in aggregate they can be used to identify and explore important variations in performance. Perhaps because of the direct connection between adverse events, patients’ healthcare experience, and the process of care itself, it can be argued that it is easier to use such information to bring about changes in organisational or clinical practice than it is with other types of information about the quality of care. Clinicians readily recognise the importance of adverse events and see the opportunities for improvement that they present.

However, the negativity of adverse events which makes them a powerful tool in quality improvement also makes it important that quality measurement does not solely focus on such events. The risk exists that a rather biased view of quality, focused on outlier events, technical quality, and patient safety issues could result.³⁸ For the clinicians involved, focusing on adverse events to the exclusion of other things could be dispiriting and demotivating.

Furthermore, there are some important practical problems involved in using adverse events in quality measurement, especially in any quantitative sense. Firstly, developing definitions of adverse events which are both mean-

ingful and can be reliably applied in measurement is difficult. Some measures rely on (and give scope for) the professional judgment of the person applying the measure, but this probably compromises reliability and may make unjustified assumptions about the knowledge and skills of that person. Other measures use more specific and detailed definitions of each type of adverse event, but these may become rigid and maximise reliability at the expense of validity. Developing definitions of adverse events which are both valid and reliable when used in measurement has proved difficult.

Secondly, most approaches to detecting and using adverse events in healthcare quality improvement make extensive use of professional review—sometimes against explicit criteria but more often on implicit basis—both to identify such events and to analyse causation and make assessments of impact and other issues. Such implicit professional reviews are often not reliable, however, and are easily biased by extraneous circumstances or information.

Thirdly, most adverse event measures are not well tested by their developers to assess and demonstrate their validity and reliability or other characteristics of their behaviour. Even where such evidence exists it appears that when measures are taken up and used by others they may not always achieve the levels of validity and reliability achieved during development, and that in particular some ongoing monitoring of reliability is needed to sustain performance.

In conclusion, although adverse events in health care provide important and useful insights into the healthcare process which can certainly be used to great effect in promoting quality and performance improvements, some

caution should be exercised, especially when they are used in measurement, either quantitatively or qualitatively.

- 1 Walshe K, Bennett J, Ingram D. Using adverse events in healthcare quality improvement: results from a British acute hospital. *International Journal of Health Care Quality Assurance* 1995;8:7-14.
- 2 Oakland JS. *Total quality management*. Oxford: Heinemann Professional Publishing, 1989.
- 3 Deming WE. *Out of the crisis*. Cambridge, MA: Massachusetts Institute of Technology, 1986.
- 4 Sanazaro PJ, Mills DH. A critique of the use of generic screening in quality assessment. *JAMA* 1991;265:1977-81.
- 5 Moser RH. Diseases of medical progress. *N Engl J Med* 1956;255:606-14.
- 6 Kane RL. Iatrogenesis: just what the doctor ordered. *J Community Health* 1980; 5:149-58.
- 7 Rutstein DD, Berenberg W, Chalmers TC, et al. Measuring the quality of medical care: a clinical method. *N Engl J Med* 1976;294:582-8.
- 8 Pringle M, Bradley CP, Carmichael CM, et al. *Significant event auditing: a study of the feasibility and potential of case-based auditing in primary medical care*. London: Royal College of General Practitioners, 1995.
- 9 Mills DH. Medical insurance feasibility study—a technical summary. *West J Med* 1978;128:360-5.
- 10 Harvard Medical Practice Study. *Patients, doctors and lawyers: medical injury, malpractice litigation and patient compensation in New York*. Harvard, New York: Harvard College, 1990.
- 11 Leape LL. Error in medicine. *JAMA* 1994;272:1851-7.
- 12 McLamb JT, Huntley RR. The hazards of hospitalisation. *South Med J* 1967;60:469-72.
- 13 Craddock JW, Bader BS. *Medical management analysis: a systematic approach to quality assurance and risk management. Volume I: introduction*. Auburn, California: Joyce W Craddock, 1983.
- 14 Wilson RM, Runciman WB, Gibberd RW. The quality in Australian healthcare study. *Med J Aust* 1995;163:458-71.
- 15 Walshe K. *The validity and reliability of adverse-event measures of quality in healthcare* [PhD thesis]. Birmingham: University of Birmingham, 1998.
- 16 Richardson FM. Peer review of medical care. *Med Care* 1972;10:29-39.
- 17 Goldman RL. The reliability of peer assessments of quality of care. *JAMA* 1992;267:958-60.
- 18 Localio AR, Weaver SL, Landis JR, et al. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. *Ann Intern Med* 1996;25: 457-64.
- 19 Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgements of appropriateness of care. *JAMA* 1991;265:1957-60.
- 20 Williamson JA, Mackay P. Incident reporting. *Med J Aust* 1991;155:340-4.
- 21 Hartwig SC, Denger SD, Schneider PJ. Severity indexed, incident report-based medication error reporting program. *American Journal of Hospital Pharmacy* 1991;48:2611-16.
- 22 Thompson JS, Prior MA. Quality assurance and mortality and morbidity conferences. *J Surg Res* 1992;52:97-100.
- 23 O'Neil AC, Petersen LA, Cook EF, et al. Physician reporting compared with medical-record review to identify adverse medical events. *Ann Intern Med* 1993;119:370-6.
- 24 Mendenhall S. The use of billing data in quality assurance. *Quality Review Bulletin* 1987;3:31-3.
- 25 Bates DW, O'Neil AC, Boyle D, et al. Potential identifiability and preventability of adverse events using information systems. *Journal of the American Medical Informatics Association* 1994;1:404-11.
- 26 Flanagan JC. The critical incident technique. *Psychol Bull* 1954;51:327-58.
- 27 Carmines EG, Zeller RA. *Reliability and validity assessment*. London: Sage Publications, 1979.
- 28 Richards T, Lurie N, Rogers WH, et al. Measuring differences between teaching and nonteaching hospitals. *Med Care* 1988;25(suppl):S1-140.
- 29 Hulka BS. Peer review in ambulatory care. *Med Care* 1979; 17(suppl):1-75.
- 30 Panniers TL, Newlander J. The adverse patient occurrences inventory: validity, reliability and implications. *Quality Review Bulletin* 1986;12:311-15.
- 31 Panniers TL. Severity of illness, quality of care and physician practice as determinants of hospital resource consumption. *Quality Review Bulletin* 1987;13:158-65.
- 32 Schumacher DN, Parker B, Kofie V, et al. Severity of illness index and the adverse patient occurrence index: a reliability study and policy implications. *Med Care* 1987;25:695-704.
- 33 Barnes C, Moynihan C. Accuracy of generic screens in identifying quality problems: analysis of false positive and false negative occurrences. *Topics in Health Record Management* 1988;9:72-80.
- 34 Brennan TA, Localio RJ, Laird N. Reliability and validity of judgements concerning adverse events suffered by hospitalised patients. *Med Care* 1989;27:1148-58.
- 35 Lohr K, editor (Institute of Medicine). *Medicare: a strategy for quality assurance*. Volume I. Washington: National Academy Press, 1990.
- 36 Brennan TA, Hebert LE, Laird NM, et al. Hospital characteristics associated with adverse events and substandard care. *JAMA* 1991;265:3265-9.
- 37 Bates DW, O'Neil AC, Petersen LA, et al. Evaluation of screening criteria for adverse events in medical patients. *Med Care* 1995;33:452-62.
- 38 O'Leary DS. Beyond generic occurrence screening. *JAMA* 1991;265:1993-4.