Strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system: literature review

By Philippe Michel

WHO commissioned Philippe Michel of the Comité de Coordination de l’Évaluation Clinique et de la Qualité en Aquitaine (CCECQA) to carry out a literature review on methods for assessing the nature and scale of harm caused by health systems. The objective of the study was to identify the strengths and weaknesses of available methods. The study, which was completed in December 2003, is a contribution to the discussion of patient safety issues.

The author alone is responsible for the views expressed in this paper. He is grateful to Jean Luc Quenon (CCECQA) who wrote the sections dealing with reporting systems, and with claims and complaints, and is indebted to Charles Vincent for his useful comments. He wishes to thank Emmanuelle Blondet, Sylvie Lascols and Vincent Mounic, of the Agence Nationale d’Accréditation et d’Évaluation en Santé (ANAES), France, for their help in the literature search and retrieval.
# Contents

**Abstract**  
**1 Introduction**  
Objective  
**2 Literature review**  
Study design  
Search strategy for articles  
Selection criteria for articles  
Definitions  
Criteria for evaluating assessment methods  
Data analysis  
Results of the literature search  
**3 Strengths and weaknesses of available methods**  
Review of medical records  
Studies based on interviews with health-care providers  
Direct observation  
Incident reporting systems  
External audit and confidential inquiries  
Studies of claims and complaints  
Information technology and electronic medical records  
Administrative data  
Autopsy reports  
Mortality and morbidity conferences  
**4 Focus on the literature from developing countries and transitional economies**  
**5 Discussion and conclusions**  
**References**
**ABSTRACT**

This report describes the strengths and weaknesses of available methods for assessing the nature and scale of harm caused by the health system, according to a defined set of criteria. These criteria, and the available methods, were identified during a WHO Working Group meeting on “Patient safety: rapid assessment methods for assessing hazards,” held in Geneva, Switzerland, 17-19 December 2002.

**Study design**

The author and his colleagues conducted an extensive search of the Medline database and other sources to identify methods for estimating hazards. Methods included: ad hoc studies based on epidemiological designs and systematic data collection (review of medical records, studies based on interviews with health care providers, direct observation); institutional and national reporting systems; external audits and confidential enquiries; analysis of existing and routinely collected data (studies of claims and complaints; information technology and electronic medical records; administrative data; autopsy reports; and mortality and morbidity conferences).

The criteria used to select methods for the present study were effectiveness in capturing: the extent of harm; availability of reliable data; suitability for large-scale or small, repeated studies; costs (financial, human resources, time and burden on system); effectiveness in influencing policy; effectiveness in influencing hospital and local safety procedures and outcomes; and synergy with other domains of quality of care.

We considered the level of evidence of published literature for rating each method on each of the criteria. When evidence-based data were available, we used a four-item scale (from + to ++++). In the absence of valid data, we presented the results in a separate table using a different rating presentation (from 1 to 4), reflecting authors’ opinions. Both ratings were considered in relation to information-rich environments. In a separate analysis, we considered the literature from developing countries. Here the small number of articles did not make it possible to extract sufficient evidence-based data, and the opinion-based rating from 1 to 4 was used.

**Results**

A total of 262 articles were selected for inclusion in the study, from 1828 retrieved references. The relevance of each method is summarized in Table 1 according to the purpose of the user, defined in terms of scope (harm, active error or latent error) and objective (adverse event counting or understanding).

<table>
<thead>
<tr>
<th>Method</th>
<th>Adverse event counting (frequency assessment)</th>
<th>Adverse event understanding (root cause analysis)</th>
<th>Latent causes / contributory factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Harm</td>
<td>Active errors</td>
<td></td>
</tr>
<tr>
<td>Review of medical records</td>
<td>X</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>Studies based on interviews with health-care providers</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Direct observation</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Incident reporting systems</td>
<td>?</td>
<td>?</td>
<td>X</td>
</tr>
<tr>
<td>External audit and confidential inquiries</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Studies of claims and complaints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information technology and electronic medical records</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative data</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autopsy Reports</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality and morbidity: conferences</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

X, method is relevant for the purpose; ?, relevance of the method for the purpose is to be confirmed.

Validity and reliability of methods
Methods based on a standardized method of data collection are more effective than other methods for estimating the nature, frequency and economic impact of adverse events. The review of medical records, the most extensively studied method, usually gives estimates that are more than ten times higher than estimates provided by reporting systems, especially for adverse drug events. The differences may be less when all adverse events are considered. In comparison with longitudinal studies based on data collected from health-care providers, the review of medical records is, however, likely to underestimate preventable adverse events, and is therefore less valid for root cause analysis. Longitudinal studies provide more reliable estimates of level of harm and may also be more reliable for estimating preventable events. Among denominator-based methods, cross-sectional studies may provide reliable estimates but appear to be less valid.

It is widely recognized that incident reporting systems cannot provide valid epidemiological data, since the number of reported incidents is likely to be an underestimate of the numerator (50% to 96% of adverse events are unreported), while the denominator (all opportunities for incidents) is unknown.

Except for electronic medical records, methods based on existing, routinely collected data are less effective than ad hoc studies. The reliability of the results using administrative data, mortality and morbidity conferences, and studies of claims and complaints may, however, be higher.

Suitability and cost

Although the feasibility of large-scale epidemiological studies is proven (numerous studies have been published), their high cost (perhaps with the exception of cross-sectional studies) and the need for extensive professional expertise prevent them from being repeated at frequent intervals. Other methods are less costly; the costs of implementing databases (electronic medical records and administrative data) were not taken into account because these methods are not set up for the purpose of assessing harm. The study of claims and complaints, along with mortality and morbidity conferences, had the best ratings on both feasibility and cost.

Effectiveness in influencing action and prevention

There are few data on effectiveness in influencing action and prevention, and the data have limited external validity because the results are likely to depend heavily on unmeasured, specific local conditions.

Focus on the literature from developing countries

The differences in ratings for methods, between developing and developed countries, may be essentially attributable to the usually data-poor environment and to the lower cost of human resources in developing countries. In developing countries, ad hoc studies based on questionnaires and interviews seem to be used more frequently than other methods.

Discussion

The available methods have widely differing purposes, strengths and limitations and must be considered as complementing each other by providing different levels of qualitative and quantitative information.

Because of the wide range of criteria considered in this review, we could not use a structured grid with explicit criteria for quality assessment of the articles. The small number of articles for each criterion and the lack of appropriate data made it impossible to perform aggregated analyses. There was a lack of evidence-based data for many of the criteria, and in those cases our appreciation was subjective. We do not claim that a simple rating of each criterion gives an adequate picture of each method. Nonetheless, the list of methodologies and the illustrative ratings, by making it easier to identify appropriate methods from among the available alternatives, should provide a starting point for national and local assessment of the nature and scale of harm, and for action on patient safety.

Data needs and future research priorities
More data are urgently needed on the reliability and validity of methods of assessment, especially regarding methods used in developing countries. These research questions should be accorded high priority by the World Alliance for Patient Safety.
1 INTRODUCTION

Many different methods are available for studying adverse events and hazards caused by the health system, and they each have strengths and limitations. Discussions of appropriate methodology in this area are frequently marred by a simplistic attempt to identify the “best” method, as if only one type of study was needed. This view is reflected by presentations and papers. For instance:

“From incident reports, the rate [of medication error] was typically about two errors per 1,000 cases. A retrospective chart review yielded a rate of seven per 1,000. A computer screening method developed at Intermountain Health Care, Salt Lake City, uncovered a rate of 38 per 1,000. Daily chart review produced a rate of 65 per 1,000. And chart review combined with computer screening revealed roughly 100 errors per 1,000 cases” (1).

This presentation highlights the importance of study design and source of information, since the measured extent of hazard clearly depends on the method used to identify hazards. However, it fails to point out that error-detection methods do not compete with each other. In the present report, we demonstrate that error-detection methods complement each other by providing different levels of qualitative and quantitative information.

While risk reduction programmes based on hazard measurement increasingly use several complementary methods, the following important points should be borne in mind.

- Rapid assessment methods are needed. Data are needed for triggering action; data collection is not an end in itself but a necessary prelude to effective action. Nevertheless, efforts to collect data should not delay action on immediate and obvious local problems.
- The relevance of the method depends on the patient safety measurement goal: a recent publication proposes a general framework to help health-care providers, researchers and administrators choose the most appropriate methods for particular goals (2).
- The relevance of a method depends on the availability of data. This review covers the results of large-scale studies in developed countries. To the author’s knowledge, there have been no national or regional attempts to assess the extent of hazards in developing countries. Access to data sources clearly differs between developed and developing countries.

The appropriateness of a method depends on the question being addressed, the resources available and the context of the study. A vast amount of opinion-based literature provides arguments for and against each method. The need to review the evidence became clear during a WHO Working Group meeting on “Patient safety: rapid assessment methods for assessing hazards”, held in Geneva, Switzerland, 17-19 December 2002 (3). The aim of the meeting was to develop guidelines on rapid assessment methods for estimating hazards. The working group identified available methods, along with criteria for assessing their strengths and weaknesses. In line with the criteria proposed during the meeting, this paper aims to develop a framework for assessing methods used in estimating hazards in health-care systems.

The methods reviewed have widely differing purposes, strengths and limitations. Obviously, a simple rating cannot give an adequate picture of each approach. Nevertheless, the list of methods and the ratings should make it easier to identify appropriate methodologies from among available alternatives, as a starting point for assessing the nature and scale of harm, and for action to improve patient safety.
Objective

The objective of the study is to describe, according to a defined set of criteria, the strengths and weaknesses of available methods of assessing the nature and scale of harm caused by the health system.

2 LITERATURE REVIEW

Study design

The present study was based on an extensive literature review.

Search strategy for articles

The author conducted an extensive search of the Medline and Embase databases, supplemented by hand searches of article bibliographies and consultations with the National Patient Safety Agency (United Kingdom) and the Agency of Healthcare Research and Quality (United States). Articles published in English and French between 1993 and 2003 were included. For articles published in other languages, the abstract (if available) was reviewed to determine whether the study findings differed from those described in the English language and French language literature.

The main Medline search strategy (search strategy No. 1) was as follows:

(Medical errors OR Medication errors OR Diagnostic errors OR Iatrogenic disease OR Malpractice OR Medical* error* [title] OR Sentinel event*[title] OR Adverse event*[title] OR Human error*[title] OR Sentinel event*[title])

AND

(Adverse drug reaction reporting systems OR Sentinel surveillance OR (Retrospective studies AND Record*) OR (Prospective studies AND Observation* [title]) OR Data collection OR Record review [title] OR Risk management OR Safety management OR Medical audit OR Audit OR Mandatory reporting OR Autopsy OR Reporting system [title] OR Morbidity mortality conference OR Mortality morbidity committee*[title])

Only titles were searched for specific expressions, as follows:

Screen* error* OR Identify* incident* OR Incident*report* OR Report* error* OR Identif* adverse effect* OR Identif*error* OR Identif* adverse event* OR Sentinel event*

In addition, a broader search was performed for the developing countries (search strategy No. 2):

(Adverse drug reaction OR reporting system OR risk management OR Safety management OR Medical audit) AND (Developing countries OR Africa OR India OR Brazil)

Selection criteria for articles

One physician (the author) reviewed the abstracts of all retrieved articles. Full text reading was done by two physicians (the author and J.L. Quenon).

We selected articles on methods for estimating hazards, including: review of medical records, studies based on interviews with health-care providers, direct observation, incident reporting systems (both institutional and national), external audit and confidential enquiries, studies of
We then assessed the methods on the basis of the following criteria:

- effectiveness in capturing the extent of harm (in different environments)
- availability of reliable data
- suitability for large-scale or small, repeated studies
- costs (financial, human resources, time and burden on system)
- effectiveness in influencing policy
- effectiveness in influencing hospital and local safety procedures and outcomes
- synergy with other domains of quality of care.

Other criteria (level of evidence and external validity) were also used, as reported below in the discussion of the criteria for evaluating assessment methods.

**Definitions**

Harm is usually measured in terms of the occurrence of adverse events. Adverse events have been defined as injuries related to medical management (in contrast to complications of disease) (4). According to the Institute of Medicine, a preventable adverse event is defined as one attributable to error (5).

Error is the failure of a planned action to be completed as intended (5). It has been defined in terms of failed processes with or without harm. The term “incident” is increasingly used for error. It is a failure in decision-making, or a failure in the process of care needed to implement good decision-making, that has the potential for causing, or that results in, an adverse event (5). There is great interest in the study of mishaps, slips, mistakes, misperceptions, near misses and other error rates, and the definition of these terms as well as the identification of a framework for errors and adverse events still call for further research (6-8). It is often claimed that prevention programmes should consider all errors, and not only focus on those directly related to harm (9-11). Case studies illustrating the usefulness of this approach are numerous (12-14). In particular, there is broad recognition that cause analysis of rare events is useful, since no epidemiological approach is possible; see, for instance, (15, 16). Studies of error rates are often based on events identified by national or local reporting systems; see, for instance, (17-19). Sometimes they may be carried out as part of clinical audit or quality assurance programmes (20, 21), mortality and morbidity conferences, or direct observation.

Experts on the management of human error have published conceptual frameworks for the cause analysis of harm and errors. Of these, James Reason’s is the most frequently used (22). In his model, he describes how errors occur despite the many existing defences and safeguards. He states that few errors occur as a result of the action of a single individual. Behind the active errors, latent human, team, task, organizational and institutional conditions cause weaknesses in the system; most error events are a result of both active error and systemic weakness (22). This framework has been adapted for health-care facilities (23-26).

The methods assessed by the present review are described in Chapter 3. Some aim to provide a quantitative assessment of the level of harm and error, while others are more relevant to carrying out qualitative analyses of causes (27). This literature review shows that methods have different strengths and weaknesses, but generally provide both quantitative and qualitative information.

**Criteria for evaluating assessment methods**

The WHO Working Group meeting on “Patient safety: rapid assessment methods for assessing hazards” held in Geneva, Switzerland, 17-19 December 2002, defined seven criteria for
evaluating methods. The first four criteria relate to the intrinsic characteristics of the methods, their validity, reliability and cost. Some of the literature directly addresses these issues. The last three criteria relate more to the ability of methods to trigger improvements in safety cultures and the quality of safety programmes (28). The seven criteria used to evaluate assessment methods are as follows.

Effectiveness in capturing the extent of harm. Comparative studies, reporting the number of events identified by one method as compared to a reference list or to another method (most often the review of medical records), provide the most evidence-based information. We also report non-comparative results. This criterion concerns the validity of methods, but absolute validity cannot be ascertained since there is no gold standard method (29). Moreover, comparisons do not really establish validity because the different methods provide different levels of qualitative and quantitative assessment of hazards. Comparisons may bring out different aspects of hazard measurement, therefore it is preferable to refer to the effectiveness of methods, rather than to their validity. The effectiveness of the review of medical records has been the most extensively studied because its use goes beyond the concerns of hazard identification: it is the usual method for peer reviews and for investigating substandard care (30). Where available, information is presented on the effectiveness of methods for studying the preventability and the underlying causes of adverse events.

Availability of reliable data. Here we report the results of inter-observer reliability studies.

Suitability for large-scale or small, repeated studies. This criterion was addressed by abstracting experiences of large-scale or small, repeated studies. The focus was essentially on acceptability and feasibility, since cost is discussed separately. Large-scale studies refer to national and regional studies. Small, repeated studies are carried out for a limited period at the hospital or local level. This criterion relates to the validity of methods and to their acceptability by health-care institutions and professionals.

Costs (financial, human resources, time and burden on system). There is no full economic evaluation of the burden related to the implementation of various methods. We therefore report the little information available on the human resources needed to implement the methods. Obtaining such information is usually a secondary objective of the published studies, and no details are available on how the calculations were performed. The financial burden obviously varies widely, depending on health-care organization and country. The most useful information comes from comparative studies. The viewpoint of the studies is usually the health-care institutions, so the organizational and operating costs (the direct costs) are measured.

Effectiveness in influencing policy. The professional literature was analysed to find instances of national, regional or local policy or strategic programmes being influenced by the publication of data collected using the methods studied.

Effectiveness in influencing hospital and local safety procedures and outcomes. Here we describe intervention studies evaluating the impact of hazard measurement on clinical practice and on improving performance.

Synergy with other domains of quality of care. Here we report experiences of the use of methods for assessing hazards in conjunction with continuous quality improvement and patient safety programmes.
Data analysis

Except for comparative studies available for the assessment of effectiveness in capturing the extent of harm, the literature consists mostly of descriptive studies. Articles on many of the topics of interest are rare, and the studies are of variable quality. We therefore included in the literature review the evidence-based articles of the highest methodological quality and did not perform any kind of aggregated analysis or meta-analysis. In addition, we considered the opinions of the authors of the articles in regard to the following: suitability for large-scale or small, repeated studies; effectiveness in influencing policy; effectiveness in influencing hospital (see Table 2) and local safety procedures and outcomes; and synergy with other domains of quality of care.

We rated each method on each of the criteria to produce a summary of its key strengths and limitations. Our approach took into account the richness of the information environment.

We considered the level of evidence of published literature. When valid information was available, we rated the criteria from + to ++++. Because of the variation in the literature, we did not use a structured grid with explicit criteria for quality assessment of the articles. A study was defined as “valid” when an appropriate description of the method (sampling strategy, data collection and data analysis) in line with current standards was available (31, 32). The lowest level (+) indicates low effectiveness or suitability or availability, or it means very high cost. The most favourable rating is ++++. Where the amount of evidence-based data is small, we note “to be confirmed”.

In the absence of valid data, we used a different rating based on authors’ opinions (see Table 3): from 1 (least favourable) to 4 (most favourable).

In a separate analysis, we considered only the literature from developing countries (see Table 4). The small number of articles did not make it possible to extract sufficient evidence-based data. We therefore used only the subjective rating scale from 1 to 4.

Results of the literature search

This report refers to 262 articles. Among the 1729 articles retrieved using search strategy No.1, we read the full text of 596 articles: 581 from the list of references generated by the computer search and 15 articles added as a result of the hand searches of article bibliographies. Of the 99 articles dealing with developing countries, retrieved using search strategy No.2, we included 75. This high proportion, compared to that of search strategy No.1, arises because we broadened the selection to include quality assurance projects that were directly or indirectly related to patient safety, such as utilization reviews or studies of referral patterns.

Figure 1   Results of the literature search
Search strategy No. 1  
(general search)

Abstract reading of 1729 references

Full text reading of 581 references

Search strategy No. 2  
(developing countries)

Abstract reading of 99 references

Full text reading of 75 references

Full text reading of 15 additional references  
(hand search from article bibliographies)

262 references  
in this report
Table 1  Overview of methods with respect to their relevance for specific purposes

<table>
<thead>
<tr>
<th>Method</th>
<th>Adverse event counting (frequency assessment)</th>
<th>Adverse event understanding (root cause analysis) latent causes / contributory factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Harm</td>
<td>Active errors</td>
</tr>
<tr>
<td>Review of medical records</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Studies based on interviews with health-care providers</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Direct observation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Incident reporting systems</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>External audit and confidential inquiries</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Studies of claims and complaints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information technology and electronic medical records</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Administrative data</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Autopsy reports</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mortality and morbidity: conferences</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

X, method is relevant for the purpose; ?, relevance of the method for the purpose is to be confirmed.

The following assessments of the strengths and weaknesses of available methods are based largely on the literature from industrialized countries (Tables 2 and 3). Developing and transitional economies are discussed in Chapter 4.
3 STRENGTHS AND WEAKNESSES OF AVAILABLE METHODS

This chapter provides explanations and caveats about the methods covered in this report.

Table 1 presents an overview of the methods discussed. We classified the methods according to their scope (harm, active error and latent error) and their objective (counting adverse events or understanding their causes). The distinction between harm and error, and the difference between looking at outcomes and seeking the systemic causes of poor outcomes, are critical in understanding the strengths and weaknesses of the methods. We also believe that it is useful to distinguish between methods likely to provide information on level of harm or active error (adverse event counting) and methods likely to provide information on contributory factors (adverse event understanding). Clearly, this classification is not absolute and many methods feature in both categories.

Review of medical records

Review of medical records is sometimes referred to as the “benchmark for estimating the extent of medical injuries occurring in hospitals”, especially as most current estimates are based on this method (33). Review of medical records is usually conducted retrospectively on a random sample of admissions. External, independent and specially trained nurses as well as medical records administrators screen inpatient records, after the index hospitalization, for various potential indicators of adverse events (e.g. death or other undesirable outcome, such as cardiac or respiratory arrest, transfer from general care to a special unit, or return to operating room). Implicit professional reviews of the records that match one or more of these indicators are then carried out by external board-certified physicians to identify adverse events, i.e. injuries attributable to medical mismanagement. For these adverse events, an assessment of negligence or of preventability is made in order to estimate the proportion of such events resulting from medical error. The method, originally developed by the California Medical Association in the late 1970s (34), was first used for epidemiological purposes in the Harvard Medical Practice Study (4), and since then has been used for almost all epidemiological studies in acute care institutions (35-41) and in other settings (42-44).

Effectiveness in capturing the extent of harm

The most evidence-based data on effectiveness come from comparative studies. Some of these studies compare the effectiveness of reviews of medical records, depending on whether the approach is implicit or explicit, and according to type of reviewer. Brennan et al. (29), before conducting the Harvard Medical Practice Study, compared reviews performed by medical record analysts and residents or fellows with reviews performed by senior physicians, considered as the reference. They concluded that reviews performed by the former were valid. A similar study comparing the accuracy of technicians and pharmacists in identifying dispensing errors reached similar conclusions (45). Weingart et al. (46) have recently raised concerns regarding the validity of reviews of medical records. They compared explicit reviews by nurses with implicit reviews by physicians and found that the numerous discrepancies between the two types of review were attributable to “nurses and physicians attending to different phenomena”, partly because, unlike nurses, “physician reviewers may not consider process problems, that are ubiquitous in hospitals, to represent substandard quality”.

Comparative studies have compared review of medical records with local voluntary incident reporting, malpractice claims and methods involving active data collection from clinical teams.

- Among 3146 admissions in a medical service of a university-affiliated teaching hospital in the United States, stimulated reporting by staff physicians, using an electronic mail
system, identified nearly the same number of adverse events as did a record review (47). However, those two methods identified only some of the patients affected; health-care providers reported significantly more preventable adverse events (62% compared with 32%, \(P = 0.003\)). Similar results were reported in an obstetrics department in the United Kingdom and in an intensive care unit in Australia (48, 49).

- If the reporting is not stimulated and if all errors are included, the ratio of reported to detected events concerning medication error is between 1 to 10 and 1 to 100 (44, 50, 51).

- Jha et al. (52) compared review of medical records with stimulated voluntary reporting and computer-based monitoring. The results included 21,964 patient-days and identified 617 adverse drug events and 86 potential adverse drug events. Of the 617 adverse drug events, the largest number was detected by review of medical records (398), followed by 275 for computer-based monitoring and 23 for voluntary reporting. There was little overlap between the events reported by the different methods: only 67 events were detected both by review of medical records and by computerized monitoring, suggesting that the actual rate is probably higher than what any single method would identify.

- Brennan et al. (53) found that reviews of risk management and litigation records missed up to 20% of adverse events, and 24% of adverse events with negligent care.

- Prospective data collection from doctors and nurses during patients’ hospital stay has recently been compared with review of medical records. The effectiveness in estimating adverse events and preventability rates, defined as the proportion of cases identified by each method compared to a reference list containing all cases confirmed by units, was calculated. The record review method was as effective as the prospective method in identifying adverse events (record review identified 70.0%, the prospective method identified 65.6%) but was less effective in identifying preventable adverse events (record review identified 39.4%, while the prospective method identified 64.8%) (54).

Comparing the results of various studies from the same environment may provide other useful data on the validity of review of medical records as an assessment method.

- In the United States, a 0.7% incidence rate of adverse drug events was estimated using review of medical records, whereas 2.4% and 6.5% rates were found, respectively, using computerized medical records and active data collection (4, 35, 55). A comparison of the Australian Incident Monitoring System (AIMS), a nationwide voluntary incident reporting system, with the results of the Quality in Australian Health Care Study based on a review of medical records showed that the therapeutic use of medications comprised the majority of adverse drug events (defined as known side-effects of a medication) identified by record review but only 8% of those identified by AIMS. Record review identified six times more adverse drug events related to problems with therapeutic use and twice as many adverse drug events related to drugs administered despite being contra-indicated, but fewer events related to non-administered medications and overdoses. Moreover, record review did not detect any event related to unintended medication administration (“wrong drug” adverse drug events) (56).

- Healey et al. (57) found that the complication rates in four surgical services, estimated using observational methods, were 2 to 4 times higher than those reported in the Institute of Medicine report (5).

Complementary points on the validity of medical record reviews include the following.

- Review of medical records is unlikely to provide valid analysis of root causes because the medical record does not include all relevant information. This is reflected in the lack of reliability of the preventability judgement. Structured methods, currently under
development, may provide more valid information on causes and methods of prevention (58).

- Studies based on review of medical records have demonstrated that the incidence of adverse events is higher among the elderly, higher in case of intra-hospital death, and increases with the length of stay in hospital (40, 41, 59, 60). These results indicate that the review of medical records has validity as a method.

- Among the potential threats to validity, hindsight bias – the tendency to impute causation when the outcome is known (61, 62) – is the most common (27, 58).

In conclusion, review of medical records based on a standardized method of data collection is effective for estimating the nature, frequency and economic impact of adverse events. Record review is nevertheless likely to underestimate preventable adverse events and specific adverse events, such as those related to drugs. Finally, it is unlikely to provide valid data for the analysis of root causes.

**Availability of reliable data**

Reliability of the results of review of medical records has two limitations: incomplete documentation in the medical record; and poor to moderate inter-observer reproducibility of the review.

Incomplete and inadequate documentation is often mentioned in the literature but has rarely been rigorously studied (63). A recent prospective observational study (64) in a surgery service of a teaching hospital concluded that only 9 out of 144 complications (6%) were not documented in the medical records, if all data sources were taken into account (mortality and morbidity rounds, and the abstract of the patient’s final medical records). The accuracy of a record should especially be questioned when the record covers a critical incident. Byrne et al. (65) showed that in the course of a single simulated anaesthesia during which there was a complex critical incident, the recording errors of 10 trainee anaesthetists increased markedly while the incident was occurring.

Reproducibility, the property of a measurement tool to yield similar results when repeatedly applied to the same phenomenon, is a major concern. The reproducibility of adverse event identification is moderate and that of preventability judgment is poor (29, 38, 40, 47, 66-69). The reliability of preventability judgement of an adverse drug event, based on explicit criteria (70, 71), may be more reliable (72).

Outcome measures seem to have higher reliability than process measures in peer reviews. Factors contributing to poorer reliability for process measures are: inability of reviewers to differentiate between cases with respect to the quality of management; bias related to the type and training of the reviewer (i.e. physician or nurse practitioner); and bias of individual reviewers (73). The poor reliability of process assessment is strongly related to difficulties in judging preventability.

Differences in medical record quality (completeness of data, validity of diagnosis, etc.) exist between countries, especially between developing and developed countries. No study was retrieved on quality of data from medical records and reliability of adverse event identification. Methods for assessing the reliability of medical records are still an issue (74). Furthermore, almost all studies were conducted in the United States, where variability in the quality of medical records may not be a concern.

**Suitability for large-scale or small, repeated studies**

The suitability of review of medical records for large-scale studies in developed countries has been demonstrated, since the largest studies in the United States and in Australia have included
thousands of medical records (4, 40, 41). However this method is time-consuming, expensive and requires extensive use of professional review (69).

International comparisons need to be interpreted carefully (75, 76). A comparative approach would require more precise definitions, as well as standardized methods of measuring and recording errors, events and, especially, preventable cases (69, 77).

Institutions have based their quality assurance programmes on data provided by review of medical records, using a prospective inclusion procedure. For example, the Royal North Shore Hospital and the Wimmera Hospital in Australia, respectively tertiary and secondary hospitals, have set up a long-term programme based on a systematic review of patient care (78, 79). The applicability of such initiatives in other settings has been discussed (80). It involves the flagging of patient records for particular problems or issues which are believed to identify a subset of patients with a suspected higher probability of quality of care problems. The initial flagging is performed by the hospital computerized system, followed by a review by trained nurse reviewers against a set of 24 general outcome criteria. Medical records with a positive screen for at least one of the criteria are further examined by the nurse reviewers and then forwarded, where appropriate, for medical review. The primary clinician reviewer will forward (to the relevant peer group) medical records that mention events which are regarded as serious breaches of the standard of care or which could reasonably be regarded as preventable.

**Costs (financial, human resources, time and burden on system)**

In a medical service of a university-affiliated teaching hospital in the United States, the cost of the review of medical records of 3146 admissions in a medical service came to US$ 54 000 (US$ 13 per admission and US$ 116 per adverse event), whereas reporting by staff physicians based on an electronic mail system during the same four-month period was US$ 15 000 (US$ 3 per admission and US$ 57 per adverse event) (81).

A prospective cohort study in nine medical and surgical units in a tertiary-care hospital over an 8-month period was performed using review of medical records, computer monitoring and voluntary reporting. The workload was, respectively, 55, 11 and 5 person-hours per week (52).

In Japan, nurses have recently compared data collection through incident reports, logs, checklists, nurse interviews, medication error questionnaires, urine leukocyte tests, patient interviews, and medical records, in order to identify adverse events in a hospital that administered both primary and secondary health-care services. The “basic method” (incident report, checklists, nurse interviews and medication error questionnaires) was found to be more effective (time consumed per ward: 20 minutes per week); patient observation and interview consumed one hour a week, while review of medical records consumed one and a half hours per week (82).

A comparison, in France, between prospective assessment by clinical teams and review of medical records concluded that the cost of the prospective method was 20% greater than that of the retrospective method (83). The mean (25% - 75% quartiles) durations of data collection per patient were precisely estimated for the 778 patients: for detection, 11 (7-12) minutes for the prospective method and 13 (5-18) minutes for the review of medical records; for confirmation, 35 (34-90) minutes for the prospective method and 18 (15-45) minutes for the review of medical records.

**Effectiveness in influencing policy**

The release of the Institute of Medicine report *To err is human* (5) generated enormous coverage in both lay and professional media (84). Although the interpretation of the results and the conclusions of the study stimulated controversy (85-88), there was, for the first time, public recognition of the burden related to adverse events.
The report was based on results of the review of medical records because no other method provided such large sets of data. A similar report was commissioned in the United Kingdom. An expert wrote that the review of medical records had undoubtedly played an important part in the development of mission statements from the Institute of Medicine in the United States and the Chief Medical Officer in the United Kingdom regarding the building of safer health systems (58).

**Effectiveness in influencing hospital and local safety procedures and outcomes**

The positive impact of the review of medical records on efforts to improve hospital practice is judged “enormous” (58). For example, the demonstration that 20% of adverse events were drug-related (4, 41) led to several studies indicating that information and decision support systems, as well as improved methods for ordering, prescribing, dispensing and administering medication, may reduce adverse events. Obviously, the review of medical records in itself has no impact on quality improvement unless it is part of a quality improvement programme (78).

Wolff and Bourke (89, 90) used a before-and-after intervention design to determine whether the review of medical records could detect and reduce adverse events occurring in emergency department attendances. They concluded that review of medical records, analysis and action can prevent recurrences. More research evaluating the impact of quality improvement actions, mostly through hospital policies and work processes, is urgently needed in the field of reduction of the risk of adverse events.

**Synergy with other domains of quality of care**

The use of adverse event monitoring, undertaken by trained audit or quality assurance staff who systematically review case records, has been advocated as a useful hospital-wide approach to clinical audit in the United Kingdom (91). Review of medical records, the cornerstone of quality assurance programmes in the United States and in many other developed countries, is thus part of quality improvement (92).
**Studies based on interviews with health-care providers**

Longitudinal and cross-sectional epidemiological studies are usually based on data collected from health professionals as the principal source of information. More rarely, information may be collected from patients. The data are collected by external investigators or using self-administered questionnaires. These methods are especially useful for the study of errors and when there is poor recording of events.

Data collection based on questionnaires and interviews is mostly used for the study of harm in the following situations:

- analysis of adverse events outside health-care institutions, including events occurring:
  - in the general population (93)
  - before admission or after discharge from hospital (94)
  - following ambulatory care (95)
  - after authorization of a drug by the administrative authority (96, 97)

- analysis of errors in health-care institutions, relating to:
  - events with poor traceability in the records, such as falls (98) or anaesthesia-related events (99)
  - latent failures (100).

Interviews are often associated with other methods (record review and reporting) for studying incidence and risk factors of adverse drug events (95, 101-103). Their use seems to be more frequent in developing countries than in developed countries.

**Effectiveness in capturing the extent of harm**

Four comparative studies consistently concluded that interview-based methods are more effective than other methods for all adverse event identification and for identifying specific risks (adverse drug events and nosocomial infections). In particular, interviews with clinical staff were found to be more effective than record reviews for identifying all adverse events in medical units, and as effective as record reviews when both the medical and surgical units were involved. In addition, interviews were more effective in identifying preventable cases (54). The relative effectiveness of interview-based methods, compared to record review and incident reporting, may vary according to specific risks (104-106).

**Availability of reliable data**

The reliability of adverse event identification using interviews with health-care providers was found to be very good (83). The reliability of cause analysis is likely to be highly variable, according to the tool, and usually poor (17).

**Suitability for large-scale or small, repeated studies**

No data were found regarding the suitability of interviews with health-care providers as a method to be used in large-scale or small, repeated studies.

According to authors’ opinions, large-scale studies seem to be suitable in developed countries: a national survey of adverse events based on interviews with health-care providers of randomly chosen units is currently under way in France. With the exception of cross-sectional questionnaire-based surveys, the suitability of this method for repeated studies remains to be demonstrated.
Costs (financial, human resources, time and burden on system)

The cost of prospective data collection through a survey based on interviews with health-care providers was 20% greater than that of record review. The cost of cross-sectional data collection was 20% lower than review of medical records and 40% lower than prospective data collection (83). When the health-care providers were interviewed by external investigators, the workload for detection and confirmation was found, on average, to be 3 hours for 25 patients. The workload was less for the retrospective method but was not considered negligible, especially when there were multiple information sources for a patient (electronic record plus notes) and when the search of the medical records was performed by the clinical secretariat (83).

Effectiveness in influencing policy

No data were found regarding the effectiveness of the interview method in influencing policy.

Effectiveness in influencing hospital and local safety procedures and outcomes

No intervention studies were found regarding the effectiveness of the interview method in influencing hospital and local safety procedures and outcomes.

According to authors’ opinions, health-care providers prefer the interview method (as compared to collecting data from medical records) because of its pedagogical and communicative virtues (54). This suggests that the interview method could be used more effectively for the continuous improvement of a hospital’s local safety and quality of care.

Synergy with other domains of quality of care

Prospective or cross-sectional data collection is frequently performed during criterion-based medical audit of practices for which there are safety concerns.
**Direct observation**

Direct observation was first used in studies of drug administration errors in the early 1960s (107). Observation is carried out by investigators either directly, for example for drug medication error studies, or using a videotape, for example for studying errors in operating theatres. Observation offers the following advantages: knowledge of error by subjects is not required, willingness to report is not required, remembering is not required, the ability to communicate is not required, and the selective perception of subjects is unrelated (108). Observer inference and effect of the observer on the observed may be more or less significant, according to the way of observing. To the author’s knowledge, this method has been used once for studying all adverse events related to surgical care (109).

Observation-based data collection is mostly used for the analysis of errors in health-care institutions, mostly active errors but also some latent errors, such as errors related to the work team, to the task, to the local environment, and to coordination between the health-care providers under observation. Direct observation is used for:

- analysis of errors, such as those occurring during preparation and administration of drugs, (110-112), and those occurring in operating theatres and during trauma resuscitations (113-115);
- assessment of safety procedures (116, 117).

**Effectiveness in capturing the extent of harm**

The performance deficiencies of airway management captured by three types of self-reports were compared with those identified through video analysis. The three types of self-reports were the anaesthesia record (a patient record constructed during the course of treatment), the anaesthesia quality assurance report (a retrospective report as part of the trauma centre's quality assurance process), and a post-trauma treatment questionnaire. Video analysis of 48 patient encounters identified 28 performance deficiencies related to airway management in 11 cases (23%). The performance deficiencies took the form of task omissions or practices that lessened the margin of patient safety. In comparison, anaesthesia quality assurance reports identified none of these performance deficiencies, anaesthesia records identified 2 (of 28), and the post-trauma treatment questionnaires suggested contributory factors and corrective measures for 5 deficiencies. Furthermore, video analysis provided information about the context of and factors contributing to the identified performance deficiencies, such as failures in adherence to standard operating procedures and in communication (118). Using a crossover study, the effectiveness of direct observation by a pharmacist was found to be greater than that of voluntary reporting for detecting medication errors (109).

The adverse event rate found by Andrews et al. (109) was higher than the rate estimated using other methods; unfortunately, this study did not use the usual definition of adverse event.

**Availability of reliable data**

No information was found regarding the availability of reliable data from direct observation.

**Suitability for large-scale or small, repeated studies**

No data were found regarding the suitability of direct observation for large-scale or small, repeated studies.

According to authors’ opinions, the direct observation method is not suitable for global assessment of adverse events. It is, however, likely to be good for focused studies, for instance...
on adverse drug events. Methods such as videotaping may be suitable for small repeated studies; their suitability for large-scale studies is likely to depend on the wide acceptance of the method by professionals.

**Costs (financial, human resources, time and burden on system)**
No data were found regarding the costs of direct observation. According to authors’ opinions, this method may be one of the most time-consuming methods.

**Effectiveness in influencing policy**
No data were found regarding the effectiveness of direct observation in influencing policy.

**Effectiveness in influencing hospital and local safety procedures and outcomes**
No intervention studies were found regarding the effectiveness of direct observation in influencing hospital and local safety procedures and outcomes.

**Synergy with other domains of quality of care**
Direct observation is frequently performed during criterion-based medical audit of practices for which there are safety concerns.
Incident reporting systems

Incident reporting is a system for detecting, reporting and analysing adverse health-care events, and learning from such events.

The scope of incident reporting systems varies greatly. For instance, the National Patient Safety Agency in the United Kingdom requires systematic reporting and analysis of near misses. Systems such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) are concerned with only a few sentinel events. A sentinel event is an unexpected occurrence that results in death or serious physical or psychological injury to a patient; it requires immediate evaluation and response.

The primary purpose of reporting is to gather qualitative data and learn from experience. Reporting is also important for monitoring progress in the prevention of errors. The system may be mandatory or voluntary. The Institute of Medicine defines mandatory reporting systems as systems run by state departments of health that require hospitals to report serious accidents and threats to patient safety. In North America, several states mandate reporting of unusual incidents and medical errors, and require all licensed health-care facilities to report these events and to propose a corrective action plan (119-121).

Incident report forms are used to document adverse and unexpected events. The reports are designed to allow characterization of the type of event and the circumstances surrounding the event in a format that can be readily entered into a computerized database. Generally, an incident report can be initiated by any member of the facility’s staff. The reports may be reviewed by the person responsible for the medical care unit. They are forwarded to the quality or risk management department.

A system for rating reported incidents is available, which scores the consequence of the event and the likelihood of recurrence (122). It also gives a clear indication of the level of investigation necessary and the response requirements.

Investigation into adverse incidents needs to be undertaken by individuals or a team who have had suitable training in clinical incident investigation. A structured investigation that aims to identify the true cause of a problem and the actions necessary to eliminate it is called root cause analysis. Comprehensive analysis of the events focuses on how and why something happened rather than who is to blame. An investigation covers the following key areas: system or organizational failings; failure in communication; failure of equipment; environmental factors; failing in the task; failure to adhere to the health and safety legal requirements; and evidence of human error (123, 124). The key to effective clinical risk management is learning from these events and disseminating the information broadly, not just locally.

Information provided by incident reporting may be particularly useful in the case of rare events, where national collection of incidents can yield a useful sample size. Moreover, a secure and confidential reporting system will allow data to be collected on a large range and number of incidents (125). Although it does not provide an estimate of the incidence of errors, a retrospective review of incident reports provides information on patterns and trends. If a sufficient number of reports are generated, these reports can identify areas in need of improvement.
Effectiveness in capturing the extent of harm

It is widely recognized that incident reporting systems cannot provide accurate epidemiological data, since the reported incidents are likely to underestimate the numerator, while the denominator (all opportunities for incidents) remains unknown (126). Stanhope et al. (49) compared an incident-reporting system with a retrospective review of 250 deliveries in two obstetric units. Staff reported only 23% of a total of 196 adverse incidents. Moreover, among the most serious incidents, only half were reported. Cullen et al. (126) prospectively investigated routine “not-stimulated” hospital incident reporting compared to observational data collection for adverse drug events in five patient care units of a tertiary hospital. Only 6% of adverse drug events were identified and only 8% of serious adverse drug events were reported. A stimulated incident reporting system based on structured confidential interviews of medical house officers by their peers identified more events and more serious events than the usual incident reporting system (127, 128).

It is estimated that 50%-96% of adverse incidents remain unreported (7, 129). Extra work, skepticism, lack of trust, fear of reprisals and the lack of effectiveness of present reporting systems are cited as key barriers (130, 131). Anonymity and medico-legal safety are important factors in encouraging staff members to report frankly on adverse events. Success of these systems is based on safe (non-punitive), simple (short incident report form) and worthwhile (dissemination of recommendations) reporting. A safety culture - rather than the current culture of blame - is needed.

Other limitations to the effectiveness of incident reporting systems in capturing the extent of harm are as follows.

- The scope is variable: some reporting systems examine the full range of adverse and unexpected events, others only record selected events.
- A reportable event may occur, but never be discovered, particularly by junior, less experienced staff (125).
- Reporting is biased: it is probable that participants are more likely to report unusual, interesting or particularly dangerous incidents (132).
- The ratings of the consequences of the event are subjective, with differences in interpretation based on individual professional experience, knowledge and values (133).

Comparison of incident rates is difficult because of differences in definitions and categorization of adverse events, and uncertainties about the case mix across institutions. Incident reporting systems are, however, likely to identify factors that may contribute to an increased risk. For example, risk factors for wrong-site surgery are: more than one surgeon involved in a case; multiple procedures on the same patient during a single trip to the operating theatre; unusual time pressures; unusual patient characteristics; and emergency cases (134).

Availability of reliable data

No information was found regarding the reliability of data produced by incident reporting systems.

According to authors’ opinions, reliability is likely to be moderate, and to vary according to individual professional experience, knowledge and values. Recommended measures to increase reliability include: clearer definitions of incidents; simplified methods of reporting; designated staff to record incidents; education, feedback and reassurance to staff about the nature and purpose of such systems; and a clear statement of the objectives of the system (131).
Suitability for large-scale or small, repeated studies

Retrospective review of incident reports usually concerns a hospital or a long-term care facility (135). Some reviews concern several departments of the same specialty, such as emergency medicine (125) and intensive care units (130). These studies cover a long period (one year or more) and include several thousand incident reports (135).

In the United States, there are four major national voluntary reporting systems: the Medication Error Reporting Program (136); MedMARx; the National Nosocomial Infection Survey; and the Sentinel Event Reporting Program. The first two programmes focus on complications of medication, and the third on hospital-acquired infections. In addition, several states have mandatory reporting systems but lack of resources limits their ability to provide better oversight and more useful feedback to hospitals. Other examples of national incident reporting systems are the Australian Incident Monitoring System (137) and the National Health Service reporting to the National Patient Safety Agency in the United Kingdom.

Incident monitoring can be successfully implemented in general practice (138). Incident monitoring systems have also been implemented in developing countries (139-142).

Costs (financial, human resources, time and burden on system)

Voluntary incident reports of adverse events have the advantage of being not as time-consuming as formal studies. Voluntary peer reporting by physicians is inexpensive and acceptable to clinician participants who spent 3-25 minutes per week, and interviewers (when reporting was stimulated) who spent 15-120 minutes per week (127).

Few estimates of costs have been reported in the literature. One study at a single centre estimated that physician reporting was less costly (US$ 15 000) than review of medical records (US$ 54 000) over a 4-month period (47). Annual funding of state reporting systems ranges from US$ 200 000 to US$ 1 500 000 (120). States may improve the efficiency of reporting if mandatory reporting is restricted to unambiguous, usually preventable, serious events, as carefully defined by the National Quality Forum (120). Another option would be an expansion of system-wide programmes, such as the Veterans Affairs programme (143) and specialty-based, focused reporting programmes, such as those for neonatal or adult intensive care units (144).

Effectiveness in influencing policy

Reporting to an external body can lead to improved safety, with alerts about new hazards and information about the experience of individual hospitals in using new methods to prevent errors. The availability of information about patient safety can assist in justifying changes or proposals requiring funding (125). For example, the Association of PeriOperative Registered Nurses developed a position statement on correct-site surgery in 2001 (134). The information can also be used as the basis for both undergraduate and postgraduate education (125).

A reporting system strengthens the message to the public that efforts are being made to understand adverse events and to reduce future risks (145).

Effectiveness in influencing hospital and local safety procedures and outcomes

Few rigorous studies have analysed the benefits of incident reporting. The only programme for which effect on safety has been demonstrated by a controlled trial is the National Nosocomial Infection Survey. Nosocomial infection rates in hospitals that implemented the full programme were 32% lower than in hospitals without the programme (146).
Interpretation of intervention studies may be difficult, since an increase in incident reporting rates may not indicate an unsafe organization, but rather reflect a shift in organizational culture towards increased acceptance of quality improvement and other organizational changes.

Recommendations for improvements may have contributed to solving problems in some areas, such as pharmaceutical care with late intravenous doses of antibiotics, narcotic errors, anticoagulant errors, and transcription errors (122), as well as in the design of medical equipment and in the development of protocols and guidelines (130). Some of these initiatives have resulted in objective improvements (122). By highlighting recurring or ongoing problems, trainees and staff will be more aware of the need to incorporate patient safety in their day-to-day activities (125).

National databases, such as the sentinel event database of the Joint Commission on Accreditation of Healthcare Organizations, contribute to the general knowledge about adverse events and lead to the reduction of risk of similar events. For example, this system has prompted the removal of concentrated potassium chloride from nursing units.

**Synergy with other domains of quality of care**

There seems to be no overlap between reporting and litigation. Several reports indicate that full disclosure reduces the risk of litigation (147).
External audit and confidential enquiries

External audit and confidential enquiries are the activities of the independent bodies, outside the health-care facilities, that are in charge of recording and analysing adverse events and other incidents. Different approaches exist. Examples of bodies conducting different types of external audit are the following:

- An independent, not-for-profit organization, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) evaluates and accredits more than 16,000 health-care organizations and programmes in the United States. In 1996, JCAHO implemented the Sentinel Event Policy and expected each hospital to voluntarily report all sentinel events and to submit a root cause analysis of these events (see http://www.jcaho.org/index.htm).

- The National Patient Safety Agency (NPSA) is a special Health Authority, created by the Department of Health in the United Kingdom. This independent body is in charge of a national mandatory reporting system. The aim of this system is to ensure that adverse events, including near misses, will be identified, recorded, analysed and reported, and that lessons learnt will be shared to effect change at local and national levels (see http://www.npsa.nhs.uk).

Both JCAHO and NPSA require health-care organizations to conduct root cause analysis and to identify improvements as part of that analysis; under certain conditions, JCAHO and NPSA may perform on-site reviews.

National confidential enquiries were initially set up in the United Kingdom. As an example, the National Confidential Enquiry into Perioperative Deaths (NCEPOD) was set up under a corporate commitment of the related Royal Colleges, Faculties and Associations, and is now an activity of the National Institute for Clinical Excellence (NICE) (see http://www.nice.org.uk). Part of the National Health Service, NICE is an independent organization responsible for providing national guidance on treatment and care for those using the National Health Service in England and Wales. The NCEPOD aims to identify potentially remediable factors in the practice of anaesthesia, surgery and other invasive medical procedures. The comments and recommendations made in the annual NCEPOD reports are based on peer review of the data, questionnaires and other records collected from hospitals, which are reviewed by NCEPOD clinical coordinators (see http://www.ncepod.org.uk/index.htm).

Other confidential inquiries from NICE cover suicide and homicide by people with mental illness, maternal deaths, and sudden infant deaths.

**Effectiveness in capturing the extent of harm**

To the author's knowledge, there is no comparative study on the extent of hazard captured by institutions or bodies in charge of external audit of adverse events or near misses.

**Availability of reliable data**

No information was found regarding the availability of reliable data. The reliability of data may be higher when the health-care providers involved in the occurrence of an adverse event also carry out the analysis. Data reliability relies heavily on the truthfulness of hospital staff, and on the ability of external reviewers to understand the nature and environment of events that occurred some weeks or months previously.
Suitability for large-scale or small, repeated studies
The examples of JCAHO, NPSA and NCEPOD, outlined above, illustrate the suitability of such approaches for large-scale studies. It seems from the literature that enquiries such as NCEPOD have also been implemented in developing countries (148).

Costs (financial, human resources, time and burden on system)
No data were found regarding costs.

Effectiveness in influencing policy
The recommendations of the bodies carrying out external audit or confidential enquiries are made easily available through publications on the respective websites (see above) and in professional journals; see, for example, (134). Examples of effectiveness in influencing policy concern wrong-site surgery and potassium chloride administration (136).

Effectiveness in influencing hospital and local safety procedures and outcomes
No data were found regarding effectiveness in influencing hospital and local safety procedures and outcomes.
According to authors’ opinions, effectiveness is likely to be low, since health-care professionals are not involved in the process.

Synergy with other domains of quality of care
No data were found regarding synergy with other domains of quality of care.
Studies of claims and complaints

A claim is defined as a demand by a patient for compensation for an alleged injury arising out of medical care (149). Complaints and lawsuits reflect patients’ subjective impressions. They are related to patients’ dissatisfaction with their physicians’ ability to establish rapport, provide access, administer care and treatment matching expectations, and communicate effectively. A small proportion (6%) of complaints become claims (150).

Studies of claims and complaints aim at establishing the legal and financial implications of adverse events. Claims give rise to various data sources, including the following: case abstracts; defence expert opinions; narrative statements from the health-care personnel involved; peer reviews; communications between the insurers’ attorneys, claims representatives and the defendant; clinical records; depositions; and legal and financial determinations (151). Experienced reviewers usually use a structured implicit review to identify the presence of an adverse event, and to evaluate its severity and preventability. The tools of root cause analysis, such as (152), are used to study latent causes of errors. Complaints are then classified into categories, such as communication, humaneness, care and treatment, access and availability, environment, and billing (153). As claims include detailed information from several sources, they may be useful for systems analysis (27).

Studies of complaints and claims also provide information regarding the causes of liability (151) and sometimes about hospital staff injuries (150).

Effectiveness in capturing the extent of harm

Comparative studies have concluded that there is a weak relationship between claims and actual adverse events or negligence (149, 154).

The probability of a claim after a negligent adverse event is 2.5% and the probability of a claim after a negligent adverse event that caused significant or major disability is 3.8% (149). Claims generally concern the most serious adverse events (either life-threatening or fatal, or costly). However, they may be made in the absence of negligence and even in the absence of any adverse event (149).

Studies of complaints and claims do not provide a denominator for determining the risk of adverse events. Moreover, assessment of error and adverse event rate is biased because specialties such as obstetrics and gynaecology, accident and emergency, trauma and orthopaedics, and general surgery are at a higher risk of claims (150).

Studies of complaints and claims have been used to assess the frequency, nature and costs of adverse drug events, and the human factor failures associated with preventable adverse drug events (151). One study estimated that 30% of claims were related to clinical negligence and could probably have been avoided (150). In an analysis of 129 adverse drug events associated with malpractice claims, adverse events were judged preventable in 73% of cases (151).

International comparisons may be hindered by confounding factors, such as the legal environment, the availability of lawyers, and the media coverage of medical incidents.

Availability of reliable data

Availability of data is a concern, since complaints and claims are handled separately and in different ways (150). Furthermore, access to data may be limited because of rules relating to confidentiality. Also, access is delayed, since the period between the date of the alleged injury and the claim may be several months or years.
Identification by two independent physicians of adverse drug events in 140 medication-related claims shows moderate reproducibility (92% agreement, kappa=0.50) (151).

**Suitability for large-scale or small, repeated studies**

Large-scale studies may be based on claims concerning insurance company malpractice records (151). Claims studies should cover a large number of physicians and practice-years, insured over a long period, ten years for example (155). In the United States, Studdert et al. (149) believe that it is difficult to have access to full information on all claims because there are several malpractice insurers. Claims studies have also been performed in large medical groups; for example, Hickson et al. (153) included 645 physicians, covering a total of 2546 years of care.

It has been claimed that these studies, if repeated, may contribute to identifying common patterns that underlie adverse events (156).

**Costs (financial, human resources, time and burden on system)**

Costs are difficult to evaluate since they arise from analysing information made available by the hospital departments dealing with complaints and claims, as well as by the external companies that provide malpractice insurance.

**Effectiveness in influencing policy**

Malpractice claims and their subsequent legal defence can entail uninsurable losses, lost practice time, damage to reputation and emotional stress (149). The primary goal of national, regional and local claim studies is to define risk management strategies to reduce the incidence of future claims and to decrease the payout in the event of a claim. An improved risk-management strategy, in particular the prompt handling of complaints, may decrease the number of complaints that become claims, and the analysis of the economic data related to claims may be helpful in planning decisions to invest resources effectively (150). Claims studies are therefore likely to strongly influence national, regional and local policies.

**Effectiveness in influencing hospital and local safety procedures and outcomes**

Malpractice claims are associated with multiple systems-related deficiencies. Analysis of complaints and claims provide a better understanding of the underlying systems and human factor failures associated with preventable adverse events. Studies of claims have made important contributions to patient safety, for instance in anaesthesia (157). Nevertheless, physicians may perceive malpractice litigation as a barrier to reducing errors and improving risk management (158, 159).

Continuing programmes to educate office-based practitioners in risk management can reduce the risk of future claims. Reductions in claims incidence for anaesthetists from 18.8% to 9.1% and in payout from 14.6% to 5%, and reductions in claims incidence for obstetrician-gynaecologists from 23.3% to 15.2% and in payout from 11.6% to 4.2%, have been reported (155).

**Synergy with other domains of quality of care**

Analysis of complaints and claims provides valuable information complementing other methods used to study adverse events, such as direct observation, voluntary reporting, interviews with physicians and nurses, and computerized approaches. It provides a basis for improving organizational structures (156).
Information technology and electronic medical records

Information technology is increasingly used to track and provide feedback about adverse events (160). Hospitals such as the Brigham and Women’s hospital, Boston, and the Latter Day Saints hospital, Salt Lake City, in the United States have pioneered the use of electronic medical records, or key pieces of information (in pharmacy and laboratory systems, for instance) for tracking and preventing errors and adverse events, and for facilitating a rapid response after an adverse event has occurred (50, 161). Recent guidelines indicate how to implement efficient systems (162).

Effectiveness in capturing the extent of harm

Information technology has been widely used to identify adverse drug events affecting inpatients. Other applications allow the detection of nosocomial infections (163) or harm associated with medical procedures such as radiotherapy (164) in inpatients, and adverse drug events or adverse events attributable to vaccination (165, 166) in outpatients. Attempts have been made to extend the search for signals (for example, readmission to hospital within 31 days, return to operating room, and surgical site infections) for all types of adverse events (167).

Classen et al. (50) combined clinical databases to detect signals that suggest the presence of an adverse drug event in hospitalized patients, such as the use of an antidote. This approach identified 81 times as many events as did spontaneous reporting. Computerized monitoring was compared with record review and voluntary reporting (52). Based on 21 964 patient-days, the study identified 617 adverse drug events and 86 potential adverse drug events. Of the 617 events, the largest number was detected by the review of medical records (398), followed by 275 by computer monitoring and 23 by voluntary reporting. There was little overlap among methods: only 67 events were detected by both the review of medical records and computerized monitoring, suggesting that the actual rate is probably higher than what any single method would identify.

In one study (165), electronic medical records were used in the outpatient setting to identify adverse drug events. The goals were to determine the frequency, types and severity of events, and compare search methods for finding adverse drug events. The following were compared: the International Classification of Disease codes; computerized event monitoring; and data mining (free-text searching for pairs of terms, such as “angiotensin-converting enzyme inhibitors” and “cough” in the same record). The study included a total of 23 062 patients, accounting for 88 154 visits. A stratified sample of the 25 056 identified incidents was assessed. The total number of adverse drug events was estimated to be 864. Almost 56% of the adverse drug events were caused by angiotensin-converting enzyme inhibitors, antimicrobials and diuretics. The sensitivity of computerized searching - compared to review of medical records as the gold standard - was 58%, with a specificity of 88% (positive predictive value 7.5%, negative predictive value 99.2%).

Compared to incident reporting, systematic computerized event monitoring identified almost twice as many patient-controlled analgesia overdoses than were reported by incident and adverse drug event reports (168).

Few studies have been reported for developing countries. Preliminary results from a tertiary hospital in Brazil have been published (169).

Availability of reliable data

No information was found regarding the availability of reliable data (74).
Suitability for large-scale or small repeated studies

The suitability of electronic records for use in studies depends on the technology available. Tracking adverse drug events does not usually require sophisticated technology. Prescriptions for antidotes or medicines used to treat adverse drug events, as well as surrogate laboratory measures (international normalized ratios, partial thromboplastin times, serum drug concentration, etc.) or International Classification of Disease codes are relatively simple to monitor and provide quantitative evidence that a medication error has occurred (170). More sophisticated systems are used in the research field but are not widespread, even in the United States (72): computerized physician order-entry systems, bar-code documentation systems, decision support tools, and such data mining tools as text searching. In 2003 in the United States, 330 health-care providers, both single-hospital organizations and multiple-facility systems, responded to a national survey of trends in information systems. Only 24% of the 330 respondents (chief executive officers, chief financial officers and chief information officers) reported that they were implementing or operating a system in which physicians entered orders and were alerted to possible adverse consequences of a drug or treatment; 40% planned to contract for such a system within the year (171). Outside the United States, some experiences have been reported (172). The absence of computerized medical data and lack of interconnectivity between systems are, however, major threats to the applicability of electronic records, especially in the developing countries.

Costs (financial, human resources, time and burden on system)

In the prospective cohort study by Jha et al. (52) in nine medical and surgical units in a tertiary-care hospital over an eight-month period, the workload was 11 person-hours per week for computerized monitoring (1 hour to generate the list and 10 hours of follow-up), 55 person-hours per week to review medical records, and 5 person-hours per week to obtain and document the voluntary reports.

The cost of software for detecting adverse drug events varies greatly. Some software is free (173). Implementing a computerized system for physician order-entry may cost an average 500-bed facility US$7.9 million in the first year and US$1.3 million each subsequent year (171).

Effectiveness in influencing policy

No information was found regarding the effectiveness of electronic records in influencing policy.

Effectiveness in influencing hospital and local safety procedures and outcomes

Information technology may be an efficient way to trigger action against errors of different types and associated adverse events. For example, Raschke et al. (174) found that 44% of alerts generated by a tool that they built had not been identified by the team of clinicians. There is some evidence that a computerized clinical information system is positively perceived by health-care providers and may reduce the occurrence of adverse events (170, 175).

Intervention studies have demonstrated the effectiveness of risk reduction programmes based on computerized event and error detection systems in routine conditions (176, 177).

Synergy with other domains of quality of care

Information technology makes it possible to develop risk management and quality assurance tools as reminders and to support clinical decisions. Such tools have proved to be amongst the most effective methods for implementing guidelines (178).
Administrative data
Algorithms have been developed to identify complications in hospital administrative databases, using - in the United States - the International Classification of Diseases, ninth revision, Clinical Modification (179-181). In 2002, the Agency for Healthcare Research and Quality developed a set of indicators, the Patient Safety Indicators, in line with a previous project, the Complication Screening Program, and on the basis of a comprehensive literature review on in-hospital complication indicators (182).

Effectiveness in capturing the extent of harm
The most complete, multifaceted validation project was carried out by the Complication Screening Program (183-185). The results are varied, suggesting that some of the complication indicators may be valid in the United States only for certain purposes, and that these indicators may be used for screening but not for measuring hospital quality and safety. In general, these indicators have a very high specificity and predictive value, and a poor to moderate sensitivity. They have been used to assess the economic burden of iatrogenic injuries (186).

Availability of reliable data
The reliability of administrative data is obviously closely linked to the completeness of the diagnosis codes. The level of completeness of these codes varies widely in the United States and in other countries (180).

Suitability for large-scale or small repeated studies
No information was found regarding the suitability of administrative data for use in studies. According to authors’ opinions, such data are likely to constitute excellent material for studies, since this source of information is presented in computer-readable form.

Costs (financial, human resources, time and burden on system)
No data were found regarding the costs associated with using administrative data. According to authors’ opinions, the data, which are recorded for other purposes, are likely to be available at minimal expense. The costs are mainly related to the development of computer algorithms.

Effectiveness in influencing policy
No information was found regarding the effectiveness of administrative data in influencing policy.

Effectiveness in influencing hospital and local safety procedures and outcomes
The release of the Patient Safety Indicators is too recent for there to be data on the impact on hospital and local safety procedures and outcomes. Only one article (187), to the author’s knowledge, refers to Patient Safety Indicators and their use in the national quality report by the Agency for Healthcare Research and Quality. It concludes that, “providers may use them to screen for preventable complication, target opportunities for improvement and benchmark performance”.

More generally, it remains unclear in the literature whether routine feedback of data on performance results in improvement in quality and safety (188, 189).

**Synergy with other domains of quality of care**

Most national performance systems include both quality and safety indicators.
Autopsy reports

*Effectiveness in capturing the extent of harm*

Autopsies are not systematically performed and reveal only one type of error, related to misdiagnosis, so this method does not provide a denominator for assessing the risk of injury. The effectiveness of autopsy reports must be considered as incremental, to complement the information provided by other methods (such as review of medical records).

The evidence-based literature on the frequency with which autopsies reveal important, clinically missed diagnoses has been reviewed (190). The 53 studies included in the review met the following criteria: the studies covered consecutive autopsies, carried out on the basis of well-defined selection criteria; the clinical diagnoses derived from the autopsy request form or from chart review; and the studies used structured classification schemes. The results of the review are as follows.

- Discrepancies between clinical diagnoses and autopsy diagnoses are not necessarily the result of medical mistakes. They also reflect situations with atypical symptoms or limited diagnostic test information.
- The base probability of an autopsy revealing a major error in a given case is 25.6%.
- The probability that autopsy will reveal a misdiagnosis that may have affected the outcome (i.e. a class I error) is 10.2% (95% CI: 6.7%-15.3%). This probability depends on:
  - autopsy rate (median 37% in the 53 studies): every 10% increase in the autopsy rate is associated with a relative decrease in class I error of 7.8%;
  - case mix (more major and class I error rates in surgery, in adults and in intensive care units) and country (slightly higher major error rate in the United States than in other countries);
  - time (1959-1999): the probability of class I error, after adjustment for autopsy rate, case mix and country, showed a relative decrease of 26.2% per decade.
- The contemporary (2000) class I error rate detected by autopsy is 3.8%-7.9%. This range reflects variations in autopsy rates from 5% to 100%.
- Clinicians cannot reliably predict which autopsies will be of high diagnostic value.

The authors conclude that the “effectiveness of autopsy for revealing unsuspected diagnoses remains sufficiently high that encouraging ongoing use of autopsy appears warranted”. The three studies conducted in India, Slovenia and the former Yugoslav Republic of Macedonia are in line with these general results.

Studies in Brazil, which were not included in the above-mentioned review, indicate similar results (191, 192).

*Availability of reliable data*

Two studies were retrieved on the availability of reliable data. The first (193) studied the inter-observer variation between pathologists in diagnosing the immediate cause of death and the principal disease involved. Two pairs of pathologists examined the organs of a total of 35 autopsies that had been performed by other colleagues as part of their daily routine. There was good agreement in the diagnosis concerning errors or principal diseases leading to death.

The second study (194) examined the inter-expert reliability of preventable death judgements for trauma. When both autopsy results and pre-hospital care reports were available, reliability of judgement between experts tended to be higher than that based only on the review of medical records.
The results of the latter study must be considered as preliminary, and further studies should be conducted based on the same method.

**Suitability for large-scale or small, repeated studies**

Autopsies have dramatically decreased in frequency, from over 40% of all non-forensic deaths in the 1960s to 6% in 1994 in the United States (190, 195) and in other western countries (196, 197). Reasons are multifactorial: changes in attitudes of clinicians and pathologists, and loss of interest in the autopsy activity (for example because of increased fear of litigation, or a communication gap between pathologists and clinicians); costs for hospitals and problems with reimbursement; some consent processes; and increasing aversion on the part of the general public to procedures such as the retention of organs (198-200). A study of junior doctors’ perceptions in England shows that they still believe the autopsy to be a useful tool, with most stating that it aims to establish an unknown cause of death (201).

Autopsy reports do not appear to be suitable for large-scale studies.

**Costs (financial, human resources, time and burden on system)**

No information was found regarding cost or other burdens of a policy of systematic autopsy.

**Effectiveness in influencing policy**

No instances were found of policies or programmes being launched on the basis of results from a series of autopsies.

The use of autopsy reports as an instrument for measuring hospital performance is hindered by problems such as the unavailability of the true denominator of interest (all patients receiving care during a given period), the frequently non-random selection of cases for autopsy, and unrecognized errors in post-mortem diagnosis (202).

In order to assess autopsy reports as a performance measure, the report of the Agency for Healthcare Research and Quality looked at studies reporting clinical diagnoses and other follow-up data on all patients admitted to a given hospital during a defined observation period, not just diagnoses for patients who died and were subject to autopsy. The report found appropriate studies for only five conditions (pulmonary embolism, acute myocardial infarction, acute appendicitis, aortic dissection and active tuberculosis) and concluded that the performance of clinical diagnosis “exhibited substantial variation, with excellent performance only for acute myocardial infarction and, to a lesser extent, pulmonary embolism” (203). The entire health-care system could therefore benefit from more autopsies as a complementary source of information on error.

**Effectiveness in influencing hospital and local safety procedures and outcomes**

No intervention study has directly addressed the impact of autopsy findings on safety practices or performance improvement.

**Synergy with other domains of quality of care**

The possible uses of autopsy in the quality assurance system and problems encountered with its implementation have been discussed (204-206). Rare studies have reported how autopsy findings from representative samples of deaths in an institution may have provided an opportunity for regular and systematic revision of medical diagnoses and treatment (207).
Mortality and morbidity conferences

Although mortality and morbidity conferences are one of the oldest quality assurance approaches in health care (208), and despite being universally familiar to physicians, they lack a precise definition, a standard format, and identified goals (209). Mortality and morbidity conferences are widespread in developed countries in surgery and anaesthesiology. They are less frequent in internal medicine and emergency medicine. Some experiences have been reported in other areas, such as radiology and psychiatry (210, 211). The goals of mortality and morbidity conferences are primarily learning from mistakes, recognizing problems, and being aware of the risks of procedures. Data are collected, complications are classified, and rates of complications may be calculated. The United States Agency for Healthcare Research and Quality has released a new form of web-based mortality and morbidity conference (212).

Commentators have, however, highlighted deficiencies in the modern practice of mortality and morbidity conferences. In particular, cases may be selected because of their educational interest, and may not involve mortality or morbidity. Thus, a mortality and morbidity conference might more appropriately be named a “case conference” (213). The widespread perception of such conferences by health-care providers in surgery is indeed that education should be the primary focus (214). The 1998 United States national survey of all internal medicine training programmes concluded that, when present, medical error is discussed with moderate to high success by 80% of the programmes, but that only half of the programmes had an established method or procedure for handling the discussion of errors. The amount of time dedicated to open discussion was highly variable, with one fifth of the programmes allotting less than ten minutes to such discussion (215). A survey of directors of emergency medicine residencies showed that most directors were satisfied with their systems but few of them rated mortality and morbidity conferences as an excellent method for the detection or prevention of clinical errors (216). Such conferences frequently serve the purpose of meeting administrative requirements, for example for quality assurance, accreditation, and other professional procedures such as verification (209, 217).

Mortality and morbidity conferences have been referred to as the “golden hour of surgical training” and have an important educational role. It is not clear whether such conferences can provide data for purposes of assessing adverse events and errors.

Effectiveness in capturing the extent of harm

Some authors advocate the epidemiological value of data from mortality and morbidity conferences (218, 219). There are, however, few comparative studies, and their conclusions are discordant. Thompson and Prior (220) found that there was excellent agreement on assessments of levels of care between those given by mortality and morbidity conferences and those arising from peer review of medical records. In contrast, a hospital prospective survey, based on daily reviews, interviews with health-care providers and observation conducted by a single observer, concluded that only 20% of the complications were presented at weekly mortality and morbidity conferences (64). Such conferences are often considered a relevant method for error analysis, despite the lack of formal demonstration of their effectiveness (27). The lack of a structured method may hamper the validity of error identification (209).

Availability of reliable data

To the author’s knowledge, there is no information on reliability of adverse event and error identification from mortality and morbidity conferences.
According to authors’ opinions, reliability is likely to be widely variable, since it depends on the objective of the mortality and morbidity conferences, the procedure for case selection, the role of the moderator, and the attendance of health-care professionals.

**Suitability for large-scale or small repeated studies**

The widespread practice of mortality and morbidity conferences attests to the suitability of their long-term implementation in hospitals, in both developed and developing countries (221, 222). Depending on local regulations, their suitability for large-scale implementation may be hampered by problems of confidentiality and medico-legal issues.

**Costs (financial, human resources, time and burden on system)**

The method of mortality and morbidity conferences is based on existing data. The cost is therefore limited and depends on the availability of computer-readable data for aggregated analysis.

**Effectiveness in influencing policy**

Mortality and morbidity conferences are often a component of hospital quality assurance policy.

**Effectiveness in influencing hospital and local safety procedures and outcomes**

A prospective survey of 332 mortality and morbidity conference case presentations conducted by trained physician observers concluded that:

"Although adverse events and errors are discussed frequently in surgical cases, teachers in both surgery and internal medicine missed opportunities to model recognition of error and to use explicit language in error discussion by acknowledging their personal experience with error” (223).

Opinion surveys and local experiences indicate that mortality and morbidity conferences play a role in influencing hospital and local safety procedures and outcomes. Nevertheless, this role may be jeopardized if the primary goal of these conferences is education, and because such conferences are frequently implemented for the purposes of meeting administrative requirements, such as quality assurance, accreditation and other professional procedures including verification (209, 217).

**Synergy with other domains of quality of care**

Mortality and morbidity conferences are usually considered as a quality assurance method (210, 216, 220, 224).

A review of the outcome of mortality and morbidity conferences in radiology found that none of the reviewed cases resulted in complaints or litigation, and no complaints or legal cases could be traced back to reported errors (225).
Table 2  Evidence-based rating of the main methods used in developed countries for estimating hazards in health-care systems (rating scale from + to ++++, the most favourable level being ++++)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ad hoc studies based on epidemiological designs and systematic data collection</th>
<th>Methods based on reporting</th>
<th>Analysis of routinely collected and existing data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Review of medical records</td>
<td>Studies based on interviews with health-care providers</td>
<td>Direct observation</td>
</tr>
<tr>
<td>Effectiveness in capturing the extent of harm</td>
<td>++++</td>
<td>++++</td>
<td>?</td>
</tr>
<tr>
<td>Availability of reliable data</td>
<td>++</td>
<td>+++ for harm assessment, to be confirmed</td>
<td>?</td>
</tr>
<tr>
<td>Suitability for large-scale studies</td>
<td>+++</td>
<td>+++ to be confirmed</td>
<td>?</td>
</tr>
<tr>
<td>Suitability for small, repeated studies</td>
<td>+++</td>
<td>+++</td>
<td>?</td>
</tr>
<tr>
<td>Costs</td>
<td>+</td>
<td>for prospective</td>
<td>?</td>
</tr>
<tr>
<td>Effectiveness in influencing policy</td>
<td>+++</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Effectiveness in influencing hospital and local safety procedures and outcomes</td>
<td>+++</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Synergy with other domains of quality of care</td>
<td>++++</td>
<td>+++</td>
<td>+++</td>
</tr>
</tbody>
</table>
Table 3  Subjective rating, where there was no evidence-based data, of the main methods used in developed countries for estimating hazards in health-care systems (rating scale from 1 to 4, the most favourable level being 4)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ad hoc studies based on epidemiological designs and systematic data collection</th>
<th>Methods based on reporting</th>
<th>Analysis of routinely collected and existing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness in capturing the extent of harm</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Availability of reliable data</td>
<td>1-2 for cause analysis</td>
<td>2-3</td>
<td>2</td>
</tr>
<tr>
<td>Suitability for large-scale studies</td>
<td>4 for harm assessment, 2-3 for cause analysis</td>
<td>2-3</td>
<td>2</td>
</tr>
<tr>
<td>Suitability for small, repeated studies</td>
<td>1 for global assessment</td>
<td>2-3</td>
<td>2</td>
</tr>
<tr>
<td>Costs</td>
<td>1</td>
<td>1-2</td>
<td>3-4</td>
</tr>
<tr>
<td>Effectiveness in influencing policy</td>
<td>3-4</td>
<td>3-4</td>
<td>3-4</td>
</tr>
<tr>
<td>Effectiveness in influencing hospital and local safety procedures and outcomes</td>
<td>2-3</td>
<td>1-2</td>
<td>3-4</td>
</tr>
<tr>
<td>Synergy with other domains of quality of care</td>
<td>2-4</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
4 Focus on the literature from developing countries and transitional economies

Literature from the developing countries was defined as the articles retrieved using search strategy No. 2 (see Chapter 2). Because of the small number of articles, we used a broader search strategy than that for developed countries, and considered all quality improvement studies that dealt with safety issues.

All types of methods were found in the literature from the developing countries. However, the number of studies based on routinely collected or existing information (169) was very low and assessments were invariably carried out in tertiary hospitals. We therefore considered that these methods were not relevant for developing countries (Table 4). The implementation of reporting systems in developing countries seems to be both rare and unevenly distributed: all reported studies were from India or Pakistan. Finally, almost all the studies specifically considered adverse drug events (139-141, 226). Only one system was implemented for all incidents (142). Direct observation (227, 228), mortality and morbidity reviews (222, 229), and autopsy-based quality control programmes (230) have been reported.

Reviews of medical records were carried out on prescription practices (231-234), assessment of mortality based on risk factors (235-237), and other processes and outcomes, such as discharges against medical advice (238) and the systematic use of Betadine cream and ointment in all burn admissions (239). Patterns of referrals and related adverse events were studied using review of medical records (240-242). Review of medical records was also used in studies that aimed to test definitions of adverse events (243). Finally, audit of practices concerning various quality and safety issues were based on record reviews (244-251). In order to guarantee the validity of the results, some authors reported that the study team supervised the registration and management of patient records at the institution concerned (242).

The number of studies based on interviews, clinical examinations and questionnaires was large, compared to the number of studies based on review of medical records. Prospective designs were mainly used, and some cross-sectional studies were reported (252). These studies dealt with safety, appropriateness and practice improvement issues. Examples of the objectives and design of questionnaire-based studies are:

- to evaluate the safety profile of ketamine when used to facilitate surgical procedures in the less controlled setting of rural hospitals in developing countries (253);
- to perform in-depth interdisciplinary internal audit of intrapartum deaths of non-malformed infants and neonatal deaths of mature infants above 34 weeks, considered both to be potentially avoidable by improved care, and in excess when compared with other regions (254);
- to analyse the contribution of adverse drug events to the overall number of visits to the medical emergency unit and to determine the proportion of events leading to hospital admissions, based on a prospective record of all visits to the medical emergency unit of an Indian hospital (255);
- to assess the time it takes from informing an anaesthetist about a case to the start of operative delivery (256);
- to set up a questionnaire on knowledge, attitude, behaviour and practice to study whether an educational intervention targeting doctors could improve the quality of care for diabetic patients (257);
to assess the appropriateness of inpatient antibiotic use, by means of a 4-week prospective survey in the medical, surgical and gynaecological wards of the Groote Schuur Hospital, South Africa (258).

Multifaceted prospective designs have also been reported, such as review followed by clinical examination (259), individual assessments of cause of death followed by regular consensus meetings (254), and criterion-based clinical audit (260).
Table 4  Rating of assessment methods used in developing countries, based on subjective appreciation (rating scale from 1 to 4, the most favourable level being 4)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ad hoc studies based on epidemiological designs and systematic data collection</th>
<th>Methods based on reporting</th>
<th>Analysis of routinely collected and existing data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Review of medical records</td>
<td>Studies based on interviews with health-care providers</td>
<td>Direct observation</td>
</tr>
<tr>
<td>Effectiveness in capturing the extent of harm</td>
<td>1-2 3-4 4 2-3 3-4</td>
<td>Availability of reliable data</td>
<td>1-2 2-3 4 for harm assessment, 2-3 for cause analysis 2</td>
</tr>
<tr>
<td>Suitability for large-scale studies</td>
<td>1 3-4 1</td>
<td>1 1</td>
<td>3-4</td>
</tr>
<tr>
<td>Suitability for small, repeated studies</td>
<td>3 3-4 2-3 3</td>
<td>3 3-4</td>
<td>3 3-4</td>
</tr>
<tr>
<td>Costs*</td>
<td>2 3-4 2-3</td>
<td>2-3 2-3</td>
<td>2-3 2-3</td>
</tr>
<tr>
<td>Effectiveness in influencing policy**</td>
<td>3-4 3-4 3-4 3-4 3-4</td>
<td>3-4 3-4 3-4 3-4 3-4</td>
<td>3-4 3-4 3-4 3-4 3-4</td>
</tr>
<tr>
<td>Effectiveness in influencing hospital and local safety procedures and outcomes</td>
<td>2-3 3-4 3-4 3-4 3</td>
<td>3 1-2</td>
<td>2-3 1-2</td>
</tr>
<tr>
<td>Synergy with other domains of quality of care</td>
<td>3-4 3</td>
<td>4</td>
<td>2-3 1-2</td>
</tr>
</tbody>
</table>

* The human cost of data collection is likely to be lower in developing countries than in developed countries.
** Effectiveness in influencing policy is rated equally for all methods because of the lack of data.
NA, the literature suggests that these methods are generally not available, except in tertiary hospitals.
5 DISCUSSION AND CONCLUSIONS

The available methods for assessing the nature and scale of harm caused by health systems have widely differing purposes, strengths and limitations, and should be considered as complementing each other by providing different levels of qualitative and quantitative information.

Of course, a simple rating based on selected criteria cannot give an adequate picture of the different approaches. Nevertheless, the list of methods and the illustrative ratings should provide a starting point for identifying appropriate methodologies among the available alternatives (Tables 2-4). Examining the capacity of these methods to provide relevant data for international comparisons was beyond the scope of this report.

Identification of appropriate methods must take into account the distinction between environments that are rich and poor as regards data, and between developed and developing countries. More generally, there is a need to consider particular national, regional and local situations. This is especially true for reporting systems and methods based on existing data. Their suitability, validity and reliability may vary widely, depending on their objectives and on the context.

Some methods, such as single case analysis and focus group discussions (261), are not assessed in this report. They may, nevertheless, be low cost and effective tools in data poor environments, for the easy identification of major problems. The use of scenarios and simulation for training and skill evaluation may provide useful information, although there are few scientific papers documenting their effects on learning and risk prevention (262).

The present review has serious limitations. The available literature varies in quality and quantity: there are few articles on large-scale studies based on interviews with health-care providers, administrative data or confidential inquiries, whereas the number of articles on reporting systems or medical record reviews is almost unlimited. Because of the wide range of criteria used in the present review, we did not include explicit criteria for quality assessment of the articles selected. The small number of articles in each category and the lack of appropriate data made it impossible to perform aggregated analyses. Furthermore, in the numerous cases where there was a lack of evidence-based data, our appreciation was subjective.

This review highlights areas for future research. In the author’s opinion, there is no need for an assessment of all criteria for each method. For example, data on costs are of limited value outside the context of the studies. Although decision-makers need data on the effectiveness of methods in influencing action and prevention at national and local level, comparative studies may have poor external validity, since the results are likely to depend on local conditions. More data are, however, needed on the reliability and validity of methods, especially methods used in developing countries. In building the World Alliance for Patient Safety, these research questions should be given high priority.
REFERENCES


163. Larsen RA et al. Improved perioperative antibiotic use and reduced surgical wound infections through use of computer decision analysis. *Infection Control and Hospital Epidemiology*, 1989, 10:316-320.


256. **O'Regan M.** Delivery times for caesarean section at Queen Elizabeth Central Hospital, Blantyre, Malawi: is a 30-minute ‘informed to start of operative delivery time' achievable? *Anaesthesia*, 2003, 58(8):756-759.


