Accountability for health outcomes and the proper unit of analysis: what do the experts think?

DAVID R. NERENZ

Center for Health System Studies, Henry Ford Health System, Detroit, MI, USA

Abstract

Background. An invitational conference was held in Dearborn, MI, in April of 1998 to discuss technical and conceptual issues related to the general topic of using outcomes data to compare plans, networks, and providers. Approximately 150 researchers, clinicians, purchasers, and representatives of accreditation bodies and government agencies attended.

Survey of participants. At the opening session, attendees participated in an electronic survey exercise designed to identify areas of agreement or disagreement on controversial issues related to the main conference topic.

Main findings. There was general agreement about the basic concept of health plan and provider accountability for health outcomes, and about the need for further development of data sources and case-mix adjustment models. There was disagreement about other issues, including questions of who should bear the cost of collecting outcomes data and whether results should be analyzed at health plan, network, or individual clinician level.

Conclusion. A group of experts agreed on the importance of reporting comparative outcomes data, but disagreed on many of the technical details of how that could best be done.

Key words: accountability, health plans, outcomes

A key feature of the Clinton administration’s reform plan for the USA’s health care in 1993–1994 was the use of standardized quality of care data by purchasing cooperatives to compare competing health insurance plans and promote the selection of high-quality health plans by individual consumers.

Even though reform legislation was not passed, many of the specific elements of the reform plan are becoming characteristics of USA health care through the aggregate effects of many smaller local initiatives. The use of report cards on health insurance plans, hospitals, or individual providers is one of those characteristics. Public and private purchasers of health care are requiring data on various aspects of performance from health insurance plans and health care providers; this information is being used in decisions concerning which plans to offer to employees on the basis of quality as well as cost [1–3]. The Health Plan Employer Data and Information Set (HEDIS) [4] from the National Committee on Quality Assurance (NCQA) is the most visible current example of such required data, but there are other examples at national, state, or local market levels. Large purchasers use the data to select those insurance plans and provider groups to be offered to employees; individual consumers use the data to select from among a (usually small) number of plans and associated provider networks offered to them during annual ‘open enrolment’ periods in which they have the option of changing insurance plans.

An emerging trend in the area of comparative performance reports is the use of outcomes data (survival, functional health status, or near-term clinical outcomes) for the purpose of selecting among competing plans and providers. The Foundation for Accountability (FACCT), a voluntary association of large public and private purchasers of health care, has been active in promoting the use of outcomes data in this way [5,6].

Controversial issues

There are both scientific and practical concerns about the use of outcomes data for the purpose of comparing plans and providers. The scientific reasons, summarized recently by Brook, Kamberg and McGlynn [7], have to do with the weak statistical associations between care processes and outcomes, and the large samples required to detect meaningful
D. R. Nerenz

Figure 1  Responses to: ‘Health plans should be held accountable for maintaining or improving functional health status in their enrollee populations’ (Item 1); ‘Large health care systems should be held accountable for maintaining or improving the health status of the residents of communities they serve’ (Item 2); and ‘Health plan, rather than provider network, is the proper locus of responsibility and accountability for health outcomes’ (Item 11).

differences in relatively rare events such as mortality or late-stage breast cancer. There are also questions about the stability of these measures over time (essential for consumer choice if one is using past performance to predict future experience) [8] and the extent to which plans or provider groups really differ significantly on many of the proposed measures [9,10]. The practical concerns have to do primarily with costs of data collection, sensitivity of existing measures, and burden imposed on patients by long and/or frequently administered surveys. In spite of these concerns, purchasers are requesting that the information be collected and reported (e.g. the inclusion of the SF-36 survey to measure ‘health of seniors’ in HEDIS 3.0).

There is also some controversy about the extent to which health plans or providers can legitimately be held accountable for the broad range of outcome measures that might be included in a measure set [11]. While there is some general consensus on providers’ responsibility for more narrowly defined outcomes such as range of motion following hip surgery or infant health following uncomplicated pregnancy and delivery, there is less agreement about other measures. Outcomes over long time intervals for chronic diseases depend a great deal on patient behavior, which is only partially under the influence of health plans or providers. Outcomes such as quality of life or return to work include dimensions that are not conceptually related to health or health care at all (e.g. financial motivation to return to work affected by level of disability insurance provided by employer or details of workers’ compensation laws). Given the number of reporting initiatives that may include some of the more controversial areas of outcome assessment, it is important to work towards a broad national consensus on those measures most likely to yield valid comparisons of plans or providers.

Agency for Health Care Policy and Research-sponsored conference

On April 22–24 1998, the Henry Ford Health System hosted a conference entitled ‘Using Outcomes Data to Compare Plans, Networks, and Providers’. The goal of the conference was to discuss the current ‘state of the art’ in the area of using outcomes data for comparative purposes, and to try to develop a stronger consensus among the parties involved about what activities within this domain are likely to have
Accountability for outcomes

Figure 2 Responses to: ‘The case-mix adjustment methods we have are adequate for making cross-plan comparisons of health outcomes’ (Item 6); and ‘The data that we need to measure health outcomes are readily available’ (Item 7).

greatest value. Approximately 150 experts on the topic from academic settings, provider organizations, health plans, corporate purchasers, regulatory and accreditation bodies, and consulting firms came together to share new findings, argue controversial issues, and identify future directions for research. Discussion focused on topics such as: consumers’ preferences for comparative data; use of risk or severity adjustment models to compare outcomes in ambulatory care settings; choice of functional status measures to assess outcomes; and new initiatives by major accreditation or regulatory bodies.

Within each session, speakers and discussion participants were asked to not only present empirical findings, but to address key questions such as the following.

- Self-reported health status as an outcome measure. Health status at a point in time, or change in health over time, using measures like the SF-36, SF-12, or the QWB scale [12] are being proposed as outcome measures for health plans and providers. This is done, in part at least, in response to plans’ and provider organizations’ claims to be responsible for the health of an enrolled population. Is this reasonable? Do we have the right tools? Do plans and provider organizations really differ on these measures? Do plans and provider organizations have sufficient control over the causal factors to be held accountable for health status outcomes?
- Risk or case-mix adjustment. There are inevitably differences among plans and provider groups in terms of characteristics of the population(s) served. Most will claim that their patients or members are ‘sicker’, but outside the inpatient settings, we have few widely accepted tools with which to evaluate that claim. HEDIS has not yet included provisions for case-mix adjustment of its measures, but the proposed FACCT measure set does. What adjustment variables should be included in required data sets? How do we decide which of several possible adjustment models are ‘correct’? How do we know when we have achieved ‘fairness’ in comparisons among competing plans or providers?

Survey of attendees during opening session

After the keynote address on the opening evening, attendees at the conference were asked to participate in an electronic survey exercise organized by Dr. Leonard Fleck of Michigan State University. The purpose of the electronic survey exercise...
was to gather information about the range of opinion among conference attendees, on key issues about the state of the field, to display the range of opinion back to the attendees in real-time and to provide the starting foundation for discussion in the specific topic session that followed.

A total of 26 questions were included in the survey (see Appendix). The questions were written specifically for the conference and the opening session, so no preliminary studies of the psychometric properties of the items were carried out. With only two exceptions, the items were phrased as statements for which a five-step response scale (strongly agree–strongly disagree) was provided.

Each attendee was given a copy of the questions. To actually conduct the survey, Dr. Fleck read each question aloud once or twice, and then gave the group 15 seconds to respond by using an electronic keypad capable of sending a signal to a computer at the front of the meeting room. Attendees could change their answer if they wished during the 15 second period by pressing a different button on the keypad. At the end of the 15 second period, the computer recorded all the responses and projected the distribution of responses on a screen at the front of the room in bar chart form. Attendees were able to see the distribution of responses for each item and comment on those responses before moving on to the next item.

Selected results
A total of 59 people actively participated in the survey. By their own self-description, 27 were researchers, 14 were administrators or managers, and the remainder were clinicians, purchasers, health policy makers, or ‘other’.

Accountability
There was general agreement among the survey participants that plans and large health care systems should be accountable for maintaining or improving functional health status, in either their enrollees (plans) or in the communities being served (large health systems) (Figure 1). Item 11, which asked the attendees to distinguish between health plan and provider network as the appropriate locus of accountability for outcomes, showed a strong preference for provider network rather than plan as the appropriate locus.
Accountability for outcomes

Technical state of the art

There was similar agreement about the lack of readily available data for making outcomes comparisons, and the lack of adequate case-mix adjustment methods (Figure 2).

Unit of analysis for variability in outcomes

Four questions asked in sequence addressed the question of whether meaningful variability in health outcomes exists at the health plan, provider network, hospital, or individual clinician levels. As indicated in Figure 3, the level of agreement rose among participants as the unit of analysis shifted from larger, more global (health plan) units to smaller units (individual clinician or hospital).

Controversial issues

There was distinct diversity of opinion on most of the remaining items in the survey. Opinion was divided, for example, on a question about whether individual consumers wanted comparative outcomes data for competing health plans (Figure 4). Opinion was similarly divided on the question of whether health care providers should bear the cost of collecting and reporting outcomes data for patients (Figure 5). Finally, although participants agreed that outcome differences among individual clinicians would be larger than those found at the level of health plan or network, there was a great deal of disagreement on the issue of which ‘unit of analysis’ (health plan, hospital, individual clinician) would yield data on differences in outcomes that would be stable from year to year (Figure 6).

Discussion

Our survey exercise has obvious limitations, as the sample was not designed to be representative of any defined population, and the items were written to provoke discussion as much as to collect analyzable data. With those limitations in mind, though, there are a few observations that can be made about the opinions of this expert group (based in the USA) on the topic of comparative outcomes analysis.

First, the idea of health plan and provider accountability for outcomes was almost universally accepted. Although there are legitimate concerns about the interpretability of outcomes differences as quality of care measures [7], attendees were very willing to endorse statements about the concept of accountability for health outcomes. The questions may therefore have more to do with the technical feasibility of collecting and analyzing comparative outcomes data than with the underlying concepts about responsibility.

Second, the group did not feel that the appropriate data elements and case-mix adjustment models for the comparative analysis of health outcomes were readily available. Inevitably, then, continued progress toward the goal of having such
data available must require additional investments in either methodological research, expansion of plan and provider data bases, or both. Public and private organizations supporting health services research (e.g. the Agency for Health Care Policy and Research in the USA) can be influential in the former area, but progress in the latter area will require an agreement by purchasers, health plans, and providers on the need to invest in more data collection and on the nature of the data to be collected. We hope that some of the other sessions in the conference, presented elsewhere in this issue, will help to provide some of the ideas needed to drive consensus on how to improve data systems.

Third, the group identified the individual clinician as the ‘unit of analysis’ at which variations in outcomes were likely to be greatest, and hospitals the unit of analysis where variations in outcome would be most stable over time. If this is actually so, then report cards comparing plans with large and overlapping provider networks will not be very useful for the purpose of consumer choice. Consumers would be expected to be much more interested in data on the performance of individual clinicians in the plan than in data on the plan as a whole. Given the limited amount of experience with consumers’ use of either plan-level or provider-level data, it is probably too early to tell what level of information will actually be the most useful. Unless plans develop methods for reducing variation among their providers, we can expect more requests for information collected and analyzed at the individual clinician level. This would have obvious implications for the size of samples required to produce reports and the kinds of health care processes and outcomes that can be studied with the individual clinician as the unit of analysis.

Finally, the lack of consensus on most items in the survey exercise beyond the fundamental questions about accountability suggests that there is still need for discussion about questions such as the appropriate unit of analysis, case-mix adjustment methods, design of health plan and provider data systems, ability of reports to adjust to changing organizational structures (e.g. mergers of health plans or high turnover among physicians in a network) and the issue of who pays for data collection and analysis.

There is clearly a great deal of research to do and much debate to take place before a general consensus can emerge on these issues. In the meantime, however, outcomes data can and should be reported and analyzed so that we have the strongest possible empirical base from which to develop consensus.

References

2. Berwick DM. The total customer relationship in health care.
Accountability for outcomes

Figure 6 Responses to: ‘In which of the four “units of analysis” are the variations likely to be largest?’; and ‘In which of the four “units of analysis” are the variations likely to be most stable over a period of 3–5 years?’ (Items 16 and 17).


Appendix

D1. Which category do you best fit in terms of professional responsibilities? 1, researcher; 2, clinician; 3, purchaser; 4, policy maker; 5, administrator/manager; 6, other.

Directions: unless otherwise indicated we will use the following response code for each of the items below: A, strongly agree; B, agree; C, uncertain; D, disagree; E, strongly disagree.

1. Health plans should be held accountable for maintaining or improving functional health status in their enrollee populations.

2. Large health care systems should be held accountable for maintaining or improving the health status of the residents of communities they serve.

3. Individual consumers want information about comparative health outcomes among competing health plans.
4. Individual consumers want information about comparative health outcomes among alternative providers and networks within health plans.

5. Health care providers should bear the costs of measuring and reporting functional health outcomes for their patients.

6. The case-mix adjustment methods we have are adequate for making cross-plan comparisons of health outcomes.

8. In 5 years case-mix adjustment methods will be adequate for making cross-plan comparisons of health outcomes.

8. The data that we need to measure health outcomes are readily available.

9. The pace of organizational change (shifting contractual arrangements) is so rapid today that outcomes data attached to plan performance 3 years ago may have little correlation with the performance of a reconfigured plan today.

10. If 25% of the staff membership of a 50-person practice changes over a 3-year period of time, then outcome performance data from 3 years ago gives us no confident basis for judging the current performance of the group.

11. It is a waste of time, money and effort to generate vast amounts of outcome data on health plans when purchasers will simply look to the bottom line (cost) to choose a plan.

12. Health plan, rather than provider network, is the proper locus of responsibility and accountability for health outcomes.

13. There is statistically significant, and meaningful variation among accredited health plans in terms of the health outcomes they produce.

14. There is statistically significant, and meaningful variation among provider networks in terms of the health outcomes they produce.

15. There is statistically significant, and meaningful variation among hospitals in terms of the health outcomes they produce.

16. There is statistically significant, and meaningful variation among individual clinicians in terms of the health outcomes they produce.

17. In which of the four units of analysis are the variations likely to be largest? a, health plan; b, provider network; c, hospital; d, individual clinician; e, don’t know or other.

18. In which of the four ‘units of analysis’ are the variations likely to be most stable over a period of 3–5 years? a, health plan; b, provider network; c, hospital; d, individual clinician; e, don’t know or other.

19. Given the research on patient non-compliance problems, outcomes data are likely to be so greatly distorted in unpredictable ways that such data will be essentially unreliable.

20. Health outcomes are generally worse in managed care settings versus fee-for-service settings.

21. For vulnerable populations (e.g. elderly, chronically ill, low-income), health outcomes are generally worse in managed care settings versus fee-for-service settings.

22. Outcomes data would be better accepted and more reliable if they were used exclusively for internal evaluation and improvement. Making such available for external use by consumers and purchasers is what threatens to corrupt the reliability of the data.

23. Getting vast amounts of audited outcome data is very expensive. This is a serious moral problem. Those funds would be better used to meet patient care needs now.

24. The outcomes movement probably missed its window of opportunity during the Clinton health reform. Since then there have been huge private investments in various outcome measurement efforts unique to individual plans and health care systems, and mostly incompatible with one another. Such sunk costs will impede (and probably defeat) any effort to establish uniform national standards for outcome measurement.

25. Outcomes data will have an impact in improving the quality of health care, or improving the quality of consumer decision making among health plans and providers, only if there is strong government regulation with respect to the generation and use of outcomes data.

26. Outcomes researchers have not paid sufficient attention to issues of privacy. Improving the ease of data collection may pose a threat to patient privacy that patients will strongly resist.

Accepted for publication 30 July 1998