The US health care industry increasingly agrees that sharing information about quality of care is necessary to stimulate providers’ efforts to improve the quality of health care. The availability of mandated, uniform clinical data in all nursing homes and home health agencies has facilitated the rapid adoption of public reporting of comparative quality data. This paper examines the conceptual and technical challenges underlying the application of information about long-term care provider quality to judge and compare the quality of care provided by nursing homes and home health agencies. In general, the impetus to apply the emerging set of quality “tools” based on mandated clinical assessments may have outstripped the evidence for their valid application in selecting top providers or for rewarding their superior performance. (J Am Med Dir Assoc 2007; 8: E.129–E.137)

Keywords: Long-term care; quality measures; nursing home; home health

Improving the quality of health care using clinical information is done either by identifying targets for quality improvement (QI) efforts or by reporting intra- or interprovider performance differences using accepted indicators of quality of care. QI can be thought of as the means of improving clinical care in specific areas whereas comparative reporting, particularly public reporting, acts as a stimulus for improvement efforts. The logic is that providers will be stimulated to invest in internal quality improvement efforts if they believe that consumers will choose providers based on public reports of provider quality or if they are otherwise rewarded or penalized based on benchmark comparisons. Both approaches require that the measurements used to characterize providers’ quality performance are reliable and valid.

Like hospitals, dialysis centers, managed care organizations, Medicare/Medicaid-certified nursing homes (NH), and home health agencies (HHA) are subject to uniform data reporting requirements. The existence of universal, mandated clinical data sets has facilitated the implementation of both internally motivated QI efforts as well as public reporting. In the case of home health agencies, the uniform clinical assessment tool mandated by the government actually grew from an impetus to have case-specific internal and external performance measures to facilitate this integrated application of quality measures. Uniform assessment mandated for nursing homes in 1991 was designed primarily for care planning but soon afterwards the data were adopted for case mix reimbursement as well as quality measurement.1–3

The purpose of this paper is to examine the conceptual and empirical validity of the data underlying the quality measurements now in use in long-term care, specifically highlighting the major assumptions the current and proposed uses these measures imply. In a call for future research, both gaps in methodological as well as applied research in this area are highlighted. Before this exposition, a brief background to quality measurement in NHs and HHAs is provided.

BACKGROUND

Nursing Home Resident Assessment

In 1984, the Institute of Medicine initiated a study of the quality of care in nursing homes. Led by Sidney Katz, the recommendations emanating from the committee’s 1986 report, Improving the Quality of Care in Nursing Homes,4 were translated almost entirely into the 1987 Nursing Home Reform Act (OBRA). One of the key recommendations of the committee was to mandate a comprehensive assessment that would provide a uniform basis for establishing a nursing home resident’s care plan—a minimum data set (MDS). The MDS was designed with input from hundreds of experts representing the academic disciplines and the professional organiza-
tions serving geriatrics, psychiatry, nursing, physical and occupational therapies, nutrition, social work, and resident rights advocates. The designers wanted an instrument that captured information needed to develop a care plan. The resident assessment instrument consists of the MDS and various resident assessment protocols (RAPs) designed to identify patients who are likely to be in need of care or treatment due to the configuration of their characteristics thought to be predictive of a given clinical problem. The RAPs are “triggered” based on patterns of responses to data elements in the MDS. An initial version was nationally implemented in late 1990, followed by a revised version introduced in 1996 that was subsequently computerized for mandatory automated submission to the Centers for Medicare and Medicaid Services (CMS) in Baltimore via states’ departments of health.

Even before a universally available patient assessment became available, data from demonstration projects were used for policy applications and not just to drive clinical care planning. The Resource Utilization Groups nursing home case-mix reimbursement system was constructed from the MDS and by 1997 became the basis for Medicare’s prospective payment system for skilled nursing facilities. Under CMS’s Nursing Home Case-Mix and Quality Demonstration, an array of quality indicators based on computerized MDS data were created. Subsequent refinements were made and a new set of indicators appropriate to the unique experiences of short-stay, postacute patients were constructed. CMS began publicly reporting comparative data, first in a 6-state pilot and then nationally in November, 2002. The revamped Quality Improvement Organizations (QIOs) funded by CMS were given the central role in working with nursing home providers to initiate quality improvement efforts. Under the “seventh scope of work” that CMS required QIOs to undertake, numerous projects in long-term care over the past 5 years have led to QIOs working with state regulatory agencies in some states to identify nursing homes needing assistance with their quality improvement activities.

Home Health Agency Outcome Measurement

Throughout the 1990s researchers at the University of Colorado worked with home health agencies to develop a system for monitoring the quality of care provided HHA patients. Based on the Outcome and Assessment Information Set (OASIS), the State of New York and The Robert Wood Johnson Foundation supported pilot testing of a quality assurance system—Outcome-Based Quality Improvement (OBQI). The OASIS data characterize patients’ diagnoses, medical condition, treatments, and functional and cognitive status. Participating HHAs reviewed reports regarding the proportion of patients who improved or deteriorated in selected domains. In 1999, CMS mandated the OASIS as a means of uniformly capturing patient level information on all home health beneficiaries. With the adoption of OASIS, the entire Medicare-certified home health care industry began to provide data to CMS required for the new Prospective Payment System (PPS) implemented in October 2000, as well as data for quality monitoring and improvement initiatives discussed in this paper.

In early 2002, CMS implemented OBQI nationally, and in April 2003, CMS launched the first phase of the Home Health Quality Initiative (HHQI) in 8 states—Florida, Massachusetts, Missouri, New Mexico, Oregon, South Carolina, West Virginia, and Wisconsin. This consumer-oriented HHQI uses a subset of 11 indicators from among the OBQI set to produce widely disseminated “report cards” on all Medicare-certified home health agencies in participating states. In 2004 this system went national and consumers are able to view agency comparisons on these QIs in local newspapers or at the CMS Web site (https://www.cms.hhs.gov/quality/hhqi/), or learn about them via telephone.

CONCEPTUAL ISSUES IN QUALITY MEASUREMENT

In addition to the technical measurement, sampling and statistical stability, and adjustment issues inherent in constructing quality performance measures, conceptual issues inherent in measuring quality are important. While some are more applicable to making comparisons between providers, others are equally applicable to interpreting the results of QI efforts. Table 1 summarizes the conceptual and technical issues associated with long-term care quality measurement.

Which Quality Measures Are Important?

The quality of long-term care is fundamentally multidimensional and encompasses clinical care issues, functional independence, quality of life and patients’ and families’ satisfaction with care. In the case of NHs and HHAs, in spite of the large array of patient level data available for quality measures, consumer advocates and various groups of clinicians do not feel that existing measures capture important aspects of quality. For example, existing NH measures do not cover psychosocial well-being or quality of life. Furthermore, although satisfaction measures are being used broadly, particularly in the NH industry, they have not been incorporated into a national reporting system. Finally, some critics of the OBQI process in home health don’t feel these measures capture the content of the educational interventions nurses provide.

Comparing Providers’ Quality

Comparing providers’ performance on quality measures essentially asks about relative effectiveness. In doing so, con-
sumers are asking whether their experience will be better in one home versus another. However, observed differences in quality measures could be a result of different providers serving different types of patients; that is, the mix of patients served differs making it difficult to determine whether observed differences are a result of serving different types of patients or because of real differences in quality. Additionally, observed differences in provider quality measures may be due to how providers go about assessing their patients, since assessment data are the basis for the measures of effectiveness. Thus, differences in measurement may undermine the validity of interfacility comparisons that are at the heart of efforts to publicly report provider performance.

**Process Versus Outcome Measures**

Quality measures can reflect providers’ performance in how treatments are administered as well as the outcomes of those treatments. The proportion of NH residents who are restrained reflects the processes of care in the home, while the proportion of HHA beneficiaries who improve in their ability to transfer from bed to chair independently is presumed to be an “outcome” of the services rendered. Establishing measures of process as benchmarks against which providers are compared assumes that there is consensus regarding what is appropriate care and which actions are inappropriate, reflecting poor care. On the other hand, outcome-based measures of quality reflect the clinically desirable result of the overall intervention; that is, of NH or home health care.

Just as in the acute care sector, there is substantial variability among the patients served and the treatments routinely provided to long-term care patients. For example, in NHs, treatments directed to postsurgical patients are very specific to wound care and recovery; improvements in walking and transfer will presumably occur naturally as a function of recovery. On the other hand, patients admitted with terminal prognoses will deteriorate regardless, although provision of specific treatments and supports could be very appropriate. The applicability of process measures is limited to those instances where there is evidence of effectiveness (eg, flu shots) or where a consensus regarding inappropriateness exists (eg, physical restraints). Presently, the CMS publicly reported NH quality measures are a mixture of process and outcomes whereas the HHA measures are virtually exclusively outcomes based.

Some have argued that both the NH and HHA quality measure models emphasize outcomes too much since important processes of care are insufficiently addressed. Measures of change in function are designed to capture the result of various rehabilitative, nursing, and medical treatments, whereas measures like pressure ulcer incidence are most often related to inadequate skin care. Nurses, whether in nursing homes or HHAs, are particularly interested in measures that can be specifically associated with the interventions they provide. Perhaps this is one reason that more global, psychosocial, quality-of-life, and well-being measures have not been successfully championed to date. At the same time, many argue that these more global measures of well-being are exactly what consumers care about, particularly in residential care.

**Validity of Quality Measures**

Assessing provider performance, particularly based on patient outcomes, implies that providers are accountable for the observed variation in the quality measure and that the quality measure is correlated with our understanding of what true quality is. Many studies have examined the relationship between staffing levels and clinical indicators of quality and others have tested whether such quality measures are correlated. Harrington and colleagues reported that nursing homes with more staff have superior quality performance, but others have not found such consistent results. Most recently, Rantz and her colleagues identified nursing homes that performed best on the CMS quality measures but medical records reviews seeking to identify good clinical care processes were found to be unrelated to quality performance. Furthermore, Schnelle and his colleagues have repeatedly found in the facilities they’ve studied that information in the records does not necessarily reflect the reality of care processes when observed by research staff.

It is important to differentiate the validity of the aggregated quality measures, used to characterize provider performance, from the validity of the patient-level data elements embedded in the MDS or OASIS. There is significant research pointing to the construct and predictive validity of MDS items ranging from cognition, diagnoses, activities of daily living (ADLs), and so forth. Similarly, several studies of the OASIS point to the validity of its data elements both in terms of the intercorrelation among pertinent items and the prediction of events such as hospitalization. However, there is far less information on the validity of the quality measures now in use, both in terms of their relationship to other structural, process, and regulatory measures of quality and in terms of whether they capture the impact of real changes in patient care activities thought to be associated with good quality. Research on the CMS NH quality measures finds they are not significantly correlated one with the other and are poorly correlated to the number or severity of regulatory deficiencies, even controlling for interstate variation in regulatory “severity.”

**Establishing Benchmarks or Comparison Groups**

Virtually all provider quality comparisons are done viz a group, or in some cases against a specific standard. Among the issues faced in establishing benchmarks are whether to use national or state benchmarks, whether to use different benchmarks for different types of providers, whether benchmarks should be fixed or should change over time, and whether benchmarks should be based on the empirical distribution observed among providers or on some external standard or guideline. There is not necessarily a “right” answer to any of these questions. For example, establishing minimums as measured by selected quality measures may not be appropriate for all quality measures since for many areas of quality performance there are no established standards that could inform where a minimum might be established. However, relying only on empirically based benchmarks (eg, below the median)
may serve to “institutionalize” the poor performance of providers operating at the median. Furthermore, while national benchmarks might make sense in the long run, large geographic differences in medical practice patterns in both acute and long-term care may mean that patients entering long-term care from hospital will have been exposed to different interventions in different areas. For example, the large inter-state variation in the use of feeding tubes among cognitively impaired residents of nursing homes is likely to affect homes’ performance on quality measures ranging from weight loss to drug use. Additionally, in markets with available alternative care options such as home health, inpatient rehabilitation, and even assisted living, the mix of patients residing in NHs will differ. Recent research on the prevalence of long-stay NH residents assessed as requiring little functional or medical support services, reveals substantial interstate variation as does the increase in case-mix acuity between 1999 and 2002.

**Technical Issues in Quality Measurement**

The following paragraphs enumerate some of the technical complications that exist in applying these measures with examples drawn from the literature or recent research. There are other technical challenges to constructing quality measures such as small sample sizes and low prevalence, and therefore the instability of the measure as well as knowing how much difference between providers being compared actually is important. However, these issues are not unique to long-term care quality measurement as much as are those issues discussed in the following paragraphs.

**Variation in Reliability of Measurement**

The reliability of the MDS and the OASIS has been extensively tested in the development and implementation phases of their use throughout the 1990s and more recently. Most data elements in the 2 instruments achieve reasonable to excellent levels of inter-rater reliability as measured by the kappa statistic. (Volume 3 of the University of Colorado report summarizing the history of the development and testing of OASIS and the OBQI process summarizes the results of several reliability studies. The investigators chose not to present the kappa statistics for low-variance OASIS items or dichotomous items with few discrepancies. Since these invariably result in lower kappa levels, slightly lower average kappa’s would have resulted.) However, most reliability tests are undertaken under optimal conditions and may not reflect “real world” conditions since providers participating in such intrusive field studies tend to differ from the average provider.

The largest multifacility reliability study undertaken to date was done as part of the NH quality measure validation study. Research nurses with established high inter-rater reliability among themselves did independent assessments of more than 5000 NH residents in 209 facilities (approximately 28 reliability assessments per facility). While high average rates of inter-rater reliability were recorded, there was substantial interfacility variation in reliability levels. Importantly, the direction of disagreement was examined and found to systematically vary both between facilities and among the facilities in the 6 states included in the study. Thus, not only were facility kappa’s systematically lower in some states, disagreements were nonrandom. In some instances facility raters were less likely to detect a problem like pressure ulcers or pain, whereas in others they were more likely than the research nurses to rate residents as having the clinical problem. Recent statistical analyses of these data reveal that directional bias in the data can result in significant differences in the relative ranking of facilities. Some in the home health and NH industry have called for more consistent training practices and facility commitment to high-quality data. The substantial interstate differences in the prevalence of some quality measures may reflect different training practices and interpretations of assessment coding conventions.

**Risk Adjustment of Quality Measures**

Risk adjustment strategies are used to try to equilibrate the patients served by the providers being compared. In addition to specifying which types of patients are included in a given quality measure (stratification), statistical regression–based approaches or stratification can be used. Stratification facilitates transparency since providers can readily identify which patients are in which strata and then focus on the highest risk patients. The regression-based approach, used in all HHA quality measures, essentially compares the observed and the expected rate of the clinical event (eg, an incident pressure ulcer) where the expected rate is predicted on that which would occur were the mix of patients in a particular provider the same as in the average provider. Both approaches have their advocates and their detractors. Stratification often results in small numbers of patients in some strata, making the resulting measure unstable. On the other hand, regression-based approaches can be very sensitive to the statistical model assumed and to its stability; eg, logistic or linear or probit.

By and large, NH quality measures, even when regression-based risk adjustment techniques are used, include fewer adjusters than is the case for the regression-adjusted models in HHA. One reason that HHAs and NHs have evolved such different strategies for risk adjustment is because many NH residents receive service for extended periods. This means that there is no relevant “baseline” variable characterizing residents’ status that might not have already been influenced by the quality of the NH. For example, being bed-bound is predictive of acquiring a pressure ulcer. However, patients may become bed-bound because of prior inadequate mobility programs. Statistically controlling for this “effect” could adjust away prior poor care. This is less of an issue in the case of HHAs since baseline measures are supposed to precede the initiation of service.

**Composite Quality Measures**

Consumers and even regulators would prefer having a single metric with which to characterize provider quality. A recent report commissioned by the Medicare Payment Commission created a single quality measure derived from the OBQI measures in order to summarize the impact of HHA PPS on home health quality (Outcome Concepts Systems,
Selection and Provider Specialization

One of the complications in comparing providers is that some types of providers attract a different mix of patients because they offer a different mix of specialty services. Considerable literature documents how hospital-based HHA or NH providers differ from those without a hospital affiliation. Additionally, the influence of specialty care units in the nursing home arena has been well-documented as a vehicle for attracting very different types of patients. Just like hospitals that specialize in open heart surgery attract the most complicated cases, specialty long-term care providers attract more complex patients from the perspective of the “risk” scoring poorly on outcome-related quality measures. Nonetheless, current practices subject all providers to the same set of quality measures. Whether this is appropriate because facilities admitting more complex cases should be able to care for them regardless or whether this excess risk has to be compensated for in the creation of the quality measures is a critical issue that plays out in the construction and interpretation of the measures.

NEEDED RESEARCH

Improving the quality of information about provider quality is one area of research where there are gaps both in the technical as well as in the conceptual. Conceptually, we need to know what consumers value and what domains of information about providers’ quality they want. Technically, we need workable models for handling measurement error that may be confounded with true quality performance and better ways of handling small samples and rare events, and measuring instability. We need to better understand the implications of establishing clinically relevant performance benchmarks, not relative to statistical averages or rankings among providers, for consumers’ and providers’ understanding and responsiveness to the information. Finally, we really need to know how responsive each quality measure is to the introduction of quality improvement interventions that are targeted to specific quality deficits considered to be related to the measure. While this question can be advanced by observing whether differences in treatment patterns and processes across providers are related to differences in the quality measures, having both observational as well as evaluative information would be invaluable. Each of these topics is discussed briefly below.

Valuing Quality of Life Versus Domains of Quality of Care

Monitoring the quality of long-term care using OASIS- or MDS-derived performance measures necessarily limits the domains of quality reported to the public. Information about quality of life, autonomy, and residents’ satisfaction is not available from either universally available instrument but some argue that it is precisely these dimensions that are of greatest concern to consumers and their advocates. Measures of quality derived from a clinical tool will necessarily be predicated on a different value base from that derived from consumers’ preferences. In the nursing home arena, the MDS 3.0, which has been under development for several years, is supposed to incorporate items asking about residents’ quality of life. Research has suggested that even demented residents can answer simple quality-of-life questions under proper conditions. Future testing of a revised MDS that includes the residents’ “voice” will have to address the fundamental issue of how to obtain unbiased satisfaction and opinion measures from residents, particularly if staff are supposed to solicit the information. Similar information could readily be obtained from recipients of HHA services, but just like NH care, it would not be appropriate to have staff ask patients about their “satisfaction” with the care they are receiving nor whether they have needs that remain unmet. To properly address these quality of life domains, substantial additional effort is needed or a completely independent source of survey information will be required.

Coping With Measurement and Statistical Complexity

Both the MDS and the OASIS have been subjected to a great deal of reliability testing. Under these volunteer “test” conditions, the instruments perform reasonably well in terms of the inter-rater reliability of the items. Nonetheless, recent research reveals that even acceptable levels of reliability still allow for systematic bias in the direction of the errors. This is consistent with evidence from analyses suggesting consistent underassessment of pain and depression. It is quite likely that a similar phenomenon would be found in HHAs with the OASIS.

Since this is likely a universal issue in facing the data on which clinical indicators of provider quality are based, generalized strategies to audit the reliability and directionality of “disagreements” are required. Furthermore, statistical models to use the results of these audits to adjust quality measures for biased measurement error are needed since it would be highly counterproductive to penalize providers whose staff conduct more thorough assessments. Based on statistical analyses of large-scale reliability data, Roy and Mor suggest a statistical model that could address this problem in conjunction with an audit. However, much more work is required to generalize this approach and it would also be important to understand how facilities’ data quality might improve were there an ongoing “threat” of data audit. Indeed, the Veterans Administration’s (VA’s) highly touted quality improvement record over the past decade is based on quality performance data that are uniformly collected by independent auditors from hospital and ambulatory care charts. While a similar wholesale approach to audited data abstraction may not be necessary in long-term care, random audits may result in greater overall data quality.
Do Improvement Efforts Translate Into Better Quality Measure Performance?

The ultimate validity of aggregated facility quality measures is predicated on their being able to detect an improvement in care processes that are known to affect an outcome being measured using the MDS. It should certainly be the case that providers instituting restraint reduction programs will see declines in the associated quality measure. Similarly, instituting rigorous pressure ulcer prevention programs, such as applying salves and turning regimes, should be reflected in both pressure ulcer incidence and prevalence. However, a variety of different clinical care interventions may contribute to reducing ADL decline rates or untreated pain, making it difficult to know whether the intervention implemented is not effective or the measure is not sufficiently sensitive.

The QIOs have been engaged in various educational and programmatic interventions with nursing homes in many states throughout the country over the past several years, trying to teach providers how to target quality improvement efforts and to systematically measure whether those targets are being met. Nonetheless, with relatively few exceptions, the results of these efforts have not been systematically described, making it difficult to disseminate the findings of these efforts in a manner that translates across areas of research.

SUMMARY

Adopting uniform, clinically relevant, patient information systems for both NHs and HHAs has already begun to transform long-term care providers. Not only do they provide the basis of a common clinical language, they form the groundwork for 2 interrelated initiatives designed to improve the care provided to long-term care patients. This impetus, probably only willingly adopted by a minority of providers in each industry, is reinforced by public reporting.

The research community and the government have a responsibility to make sure that the technical aspects of the quality measures being used to compare NH and HHA providers are up to the challenge of being used both to stimulate the organizational changes needed to redesign care processes to improve care and to allow for legitimate and valid comparisons across providers. The current crop of measures, while a major advance over the limited validity of the admittedly idiosyncratic survey and certification process, continue to leave much to be desired. While they appear to be reliably measuring quality in select areas, the validity of the measures in terms of capturing a global notion of quality is questionable. Furthermore, the inconsistency of measurement across providers may undermine the legitimacy of the comparisons for which these measures were created. There is evidence that this is the case in NHs, whereas in the case of home health care the research has not even begun. Nonetheless, we should not stop public reporting or other uses of quality measures simply because they continue to have significant deficits.

There are approaches that can be instituted that have a high likelihood of improving the data and therefore the validity of provider comparisons. Even though CMS is in the process of introducing demonstration projects focused on paying NH providers who perform well on the current crop of quality measures, it is probably premature to be doing so, particularly since all but the most sophisticated pay for performance systems that are likely to be based on a composite measure that is known to mask differences between facilities by combining uncorrelated measures.

REFERENCES


37. Feng ZG, Intigrator D, Mor V. The effect of state case-mix payment on nursing home resident acuity. Health Serv Res 2006;41:1317–1336.


44. Mukamel DB, Spector WD. Nursing home costs and risk-adjusted outcome measures of quality. Med Care 2000;38:78–89.


**DISCUSSANT**

Dan Berlowitz, MD, MPH, Center for Health Quality, Outcomes and Economic Research

I have always been a strong believer in the value of quality indicators derived from databases such as the MDS or OASIS (although as Dr Schnelle is to describe, perhaps I shouldn’t be so enthusiastic). We know that there are serious problems with the quality of long-term care, as there are with most other areas of health care. And one of the basic principles of quality improvement is very simple, if you can’t measure it, you can’t improve it. We need reliable and valid quality indicators if we are to make improvements in the delivery of long-term care. While I could speak for hours about the strengths and weaknesses of Quality Indicators, I must go back to the theme of this conference, “Physician Practice in Long-Term Care.” What is the relevance of such quality indicators to physician practice? As I am not sure I can entirely answer this question at the outset, I will instead pose several related questions that will illuminate some of the issues.

First, to what extent do these quality indicators capture the quality of physician care in long-term care settings? Certainly the appropriate actions to prevent and treat many of the geriatric syndromes encompassed by these quality indicators require a multidisciplinary team-based approach to care. The physician is a central member of this team. Yet at the same time, I am concerned that the individual physician’s ability to influence these quality indicators is often quite limited. The clinical condition that I have extensively thought about is pressure ulcers. When I lecture physicians on this topic, I always say that no matter how good a clinician you are, you alone cannot prevent most pressure ulcers. No matter how many orders you write to turn a patient every 2 hours, you cannot ensure that this important action takes place. Rather, the long-term care delivery system must be organized and oriented to ensure the provision of appropriate and consistent care. Creating an organization that can do this is the real
challenge. Thus, while quality indicators may provide useful information about quality of care for clearly important conditions, they really capture how the team is functioning and probably provide limited insight into the practices of the individual physician. High-quality physician care is often necessary to perform well on these quality indicators but is hardly sufficient.

Second, what should the individual long-term care physician do when confronted with consistently poor performance on these quality indicators? There may be some short-term fixes that the physician may be able to implement, but I really believe what is usually required is a substantial change in the underlying culture and organization of the health care delivery team. Such fixes are never easy and require considerable resources, effort, and a true desire to implement change. The physician may be able to promote such changes, but it will require a true commitment also from other leaders of the care team. And it is important to recognize that there is very little incentive for a physician to take part in these quality improvement and reorganization efforts. He certainly will not see increased reimbursements as a result of efforts to improve care. Thus, I believe the physician is put in a very difficult position when faced with consistently poor performance. Do I continue to do the best I can in the circumstances or do I try to alter the circumstances with only limited chances for success?

My third and final question is whether we even have the tools with which to assess the quality of physician care in long-term care settings? Here, I believe a start has been made in process measures assessing care. One study reported the development of flow charts that described potential paths of care for 3 common events: fever, chest pain, or shortness of breath. The flow charts were then combined with chart abstractions to create clinical scenarios describing the actual process of care. Physicians then rated these scenarios on specific subscales with judgments of quality made based on implicit criteria. More recently, a significant advance in process measurement was achieved through the Assessing Care of Vulnerable Elders, or ACOVE, project. Through a process that combined literature reviews, examination of guidelines, and expert panels using a modified Delphi process, consensus was achieved on a set of quality indicators that covered areas such as medication use, end-of-life care, and management of specific conditions such as hypertension and diabetes. While the ACOVE quality indicators were originally developed for community-living persons at high risk for death or functional decline, Dr Saliba developed a modified set of indicators specific to NHs. Some of the measures are very similar to HEDIS measures being used for physicians in managed care settings. For example, if a NH resident has diabetes, then his or her glycosylated hemoglobin levels should be measured at least every 12 months or care goals or other records should indicate why this is not appropriate. These quality measures very much focus on the effectiveness of care. Other important dimensions of quality are not captured. Is physician care timely; this is an area in which there are significant concerns. What about whether the care is patient centered? Quality measures for these areas still need to be developed. Moreover, even for those areas where we have some preliminary measures, such as the ACOVE indicators, significant questions remain. How reliable and valid are these data? How can they be used in profiling care? While this initial work is promising, further research is required before we can accurately measure the quality of physician care in long-term care settings.

What then can I conclude about Dr Mor’s presentation? Quality indicators are a significant advance that should do much to improve care for long-term care patients. But if we are concerned, or want to improve the quality of physician care, they may not be enough. As with other health care settings, data on physician performance will be desirable and useful, there just is a long way to go in long-term care.

DISCUSSANT

John F. Schnelle, PhD, Departments of Medicine and Public Health, University of California

The adoption of the MDS represents a major achievement in long-term care, and issues concerning MDS accuracy or its association with validated assessment scales should not be the primary concern of a medical director. Whatever the MDS accuracy or validity problems may be, they can certainly be corrected if the causes for the inaccuracy are identified.

In this regard, it is important to distinguish 2 types of paradigms to assess accuracy. In one paradigm research nurses complete MDS or standardized assessments using at least some information gathered through indigenous nurse interviews or medical records. This paradigm checks the degree to which NH staff report information to research nurses that is consistent with how other or perhaps even the same NH staff report this information on the MDS. A second paradigm requires research staff to assess residents independently. For example, the accuracy of MDS reports of pain, depression, and ADL dependence have been assessed by directly administering standardized assessments to residents. The standardized assessments are then compared to what is reported on the MDS by indigenous nurses.

Investigators using the latter type of accuracy paradigm have generally reported more accuracy problems than investigators using other methodologies. The systematic errors reported (pain and depression are underreported and ADL dependence is consistently overreported) are no doubt a result of the failure of NH staff to implement validated assessments when completing the MDS. For example, if NH staff do not ask residents about pain, then they will not report pain symptoms on the MDS and will furthermore tell research nurses during “independent” interviews that the residents do not have pain. An understanding of these systematic errors suggests interventions that should improve MDS accuracy.

The much bigger issue to medical directors is how to use the MDS in improvement programs. If the outcomes measured by the MDS are more influenced by factors other than the processes of care under the control of providers, then there is the potential to be misled. One might interpret either random or systematic changes in the MDS data due to changes in resident acuity as reflecting a change in care. Making such erroneous interpretations can lead one to con-
clude that ineffectual improvement interventions are success-
ful and that effective ones are not.

If providers lack accurate and timely measures of the daily
processes of care that residents receive, then the probability of
making serious mistakes about quality issues increases when
monitoring outcomes. These points are well addressed in the
continuous improvement literature, which provides models
for many of the improvement efforts in NHs today. These
continuous improvement models were originally based on
statistical process control (SPC) technology because of the
central importance of process measurement in the improve-
ment system. In these models, the measurement of processes
are the first step to improvement.

Unfortunately, many in the provider and regulatory com-
munity may believe that monitoring MDS data and providing
feedback or “in-service training” to direct care staff when
changes or high prevalence rates of poor clinical outcomes are
noted is an adequate method to improve care. The argument
is often made that a “good leader” in the NH will know how
to use this information to improve MDS outcomes even if it
is unknown if there are adequate information and staffing
resources to improve actual care delivery. Presumably, if one
is considered a “good leader,” one might not even convey to
upper level administration that there are resource problems
that preclude good care.

This is an unfortunate management strategy, which is un-
fair to direct care providers including medical directors. If the
MDS QIs are not accompanied by equally formal and accurate
methods to measure direct care process delivery, then care will
not improve. Furthermore, significant resources will be spent
superstitiously reacting to changes in outcomes that have
nothing to do with care process delivery.

“Good leaders” in the NH, including medical directors,
should insist that an objective analysis be completed of the
resources necessary to provide care that might influence the
MDS outcomes and evidence about how much even optimal
care can influence some of the outcomes. This resource anal-
ysis should emphasize not only the staffing requirements nec-
essary to provide “good care” but also the training and infor-
mation systems necessary to manage care.