The Cost of End-of-Life Care
A New Efficiency Measure Falls Short of AHA/ACC Standards

Gerald W. Neuberg, MD, FACC

Skyrocketing costs jeopardize health coverage at every level and impede efforts to cover the uninsured. Observations that higher cost care is unassociated with improved outcomes have further stimulated efforts to gain more value for our healthcare dollars. To identify and deter unnecessary care (estimated to account for ≈30% of all health costs), new provider performance measures and incentives are being developed, but we must understand their strengths and weaknesses to ensure that they do not misrepresent or compromise care.

Care near the end of life consumes a disproportionate share of costs and is a logical target for efforts to promote value in health care. Recently, Consumer Reports launched a new online tool that rates the relative “aggressiveness” (or inefficiency) of US hospitals, based on the average intensity and cost of inpatient care in the last 2 years of life of chronically ill Medicare beneficiaries, as reported in the 2008 Dartmouth Atlas. This new measure is the primary focus of this review, which includes an assessment according to the recently published American Heart Association (AHA) and American College of Cardiology (ACC) standards for public reporting of efficiency measures. Alternate measures including risk-adjusted cost savings and avoidance of nonrecommended care are also discussed and compared.

The 2008 Dartmouth Atlas
Wennberg et al retrospectively measured inpatient costs over the last 2 years of life in Medicare beneficiaries with serious chronic illnesses—including heart failure (CHF), lung disease, cancer, dementia, vascular disease, kidney disease, and liver disease—who died between 2001 and 2005 after receiving care at nearly 3000 US hospitals. They found wide variations in end-of-life spending, even after adjusting for age, gender, race, and type of chronic illness. Among top teaching hospitals, Mayo Clinic had the lowest cost at $53 432 per patient, whereas the University of California, Los Angeles and New York University had the highest costs at $93 000 and $105 000 per patient, respectively. High-cost centers provided more services including more hospital and ICU days, doctor visits, and specialist consultations.

Because all patients had the same fatal outcome, the authors believe that the extra care provided at high-cost centers was unrelated to illness severity and therefore unwarranted. On the basis of prior work, they also believe that such care generally is physician mandated and supply and profit driven such that, if all physicians were trained and incentivized to practice as efficiently as the salaried Mayo physicians, billions of dollars would be saved without sacrificing quality or outcome. They further recommend their model be used to rate provider behavior, allowing patients and payers to favor providers whose practice patterns they prefer.

This dramatic report has received considerable attention in the press and in Washington. The Congressional Budget Office wondered, “How can the best medical care in the world cost twice as much as the best medical care in the world?” The data do raise many important questions, and such large variations are intuitively disturbing, but the study also harbors methodologic issues that are easily overlooked by a nonclinical audience.

Disease Severity and Prognosis in Retrospect
Normally, retrospective studies comparing treatment outcome and cost among patient groups must adjust for differences in disease severity. Intensity of care is a well-known severity marker, such that patients spending more days in hospitals and intensive-care units are expected to appear sicker at comparable time points than patients requiring less care. Even after correcting for such differences, residual confounding by indication still can cause underestimation of treatment benefit or exaggeration of harm. In the current study, Wennberg et al did not measure or adjust for severity, as they believe their model involves measures of provider “efficiency and performance that minimize the chance that variations in the care can be explained by differences in the severity of patients’ illnesses.” They further state that “by looking at care delivered during fixed intervals of time before death, we can say with assurance that the prognosis of all the patients in the cohort is identical—all were dead after the interval of observation.”

From a clinical perspective, this retrospective logic misrepresents the prognostic and therapeutic uncertainty that we must contend with in real time. What matters in providing care are the apparent severity and treatability of illness at the time of patient evaluation, not at the time of death. Thus, the
faintest way to assess treatment efficacy and efficiency is to assemble cohorts with comparable disease burdens at time zero, and then track subsequent outcome and resource utilization in survivors and decedents. In contrast, looking back at fixed intervals before death identifies patients whose condition at time zero varies markedly, more so for longer intervals, and this alone could explain substantial variation in resource allocation. Furthermore, end-of-life spending does not reveal whether a provider’s efforts effectively saved, extended, or improved any lives. For example, end-of-life costs cannot distinguish a patient who lives 24 months (on whatever treatment) from a sicker patient who would have lived 12 months on the same regimen, but instead survives 24 months with more aggressive care. From the look-back perspective, care is viewed not as a means to improve health, but as an accumulation of expenses that failed to prevent an inevitable death.

End-of-life spending would be a more straightforward indicator of provider performance if diseases progressed and presented in a uniform fashion, but this is not the case. In patients with fatal CHF, at least one third die unexpectedly, whereas most others experience progressive CHF requiring episodic hospital treatment before their demise. By the authors’ method, if my practice randomly sees a greater proportion of inexpensive sudden deaths, we will be rated undeservingly as more efficient than others who see a higher rate of costly progressive CHF. However, if we prevent sudden deaths by implanting more defibrillators, we will see and treat more progressive CHF (because of the competing risks of these outcomes), and our efficiency rating will decline. If we offer such patients greater access to life-extending procedures like biventricular pacing or cardiac transplantation, our rating will plummet further, because they are sick enough that some will not survive beyond the measured interval after costly treatment, regardless of how appropriately or expeditiously it was provided.

Many Reasons for Care

Was all cost variability in the Dartmouth Atlas due to varying severity and course of illness? Of course not. The University of California, Los Angeles and the New York University patients are not twice as sick as Mayo’s. Did unnecessary care account for some of the results? Yes, undoubtedly, but the Medicare database does not tell us how much care was unwarranted, or why it was provided. There are many causes of variability in care, including what might be called social care, defensive care, desperation care, and limbo care, none of which is strictly physician mandated or profit driven.

Social care refers to extra hospital days accrued by patients whose medical problems would be manageable at home if they had better compliance, follow-up, family support, or home care coverage, and by those who require nursing home placement. In such cases, hospital discharge delays are common for myriad reasons having little to do with provider performance. In one study of medical inpatients, 17% of all hospital days were classified as medically unnecessary “delay” days, and the most important cause was unavailability of postdischarge facilities. Days spent awaiting postdischarge facilities (primarily nursing homes) represented 41% of all delay days.

Defensive care refers to interventions directed at avoiding lawsuits, such as the routine hospitalization of patients presenting to an emergency room with chest pain, and imaging procedures that are done for any symptom that is remotely suggestive of cancer. A study showing lower health costs in states with versus without recent tort reform (mainly caps on awards) confirmed that malpractice pressure is associated with defensive care. The effect was modest—under 10%—but the magnitude of defensive care was probably underestimated, because physicians worry more about the personal costs of becoming tangled in litigation than about the size of malpractice awards, for which they are insured.

Desperation care is given to dying patients whose families cannot accept that illness has become irreversible and that continued aggressive care will only cause suffering. Such families may keep insisting that “everything” be done to keep their relative alive, long after clinicians see no possibility of recovery. Sadly, too many patients die in isolation, pain, and fear in our hospitals and intensive-care units, and palliative care is underutilized. Yet, in the experience of medical ethics consultants, patients or families wishing inappropriate end-of-life care far outnumber cases of doctor-mandated inappropriate care. Barriers to compassionate end-of-life care include unrealistic expectations of cure, family psychodynamics, distrust of the system, and other cultural or religious factors, compounded by poor planning and communication.

Ironically, since modern medicine can stabilize so many chronic, acute, and critical illnesses, even in the elderly, it can be difficult to tell when the end of life begins. Such uncertainty is typical of advanced cardiopulmonary disease, whose course is punctuated by frequent but often reversible decompensation. Advanced cancer and dementia progress more predictably, but still may be interrupted by treatable events like infections. When there is no reasonable hope of recovery, everyone may readily agree to withhold aggressive measures and focus on the patient’s comfort and dignity. But when the prognosis is uncertain, the main problem is not whether to initiate acute care, but when and how to stop if treatment fails. Too often, such patients lose capacity without leaving instructions, and difficult decisions fall on unprepared relatives.

Advance planning is a shared responsibility. Some question the utility of advance directives, but I consider them very helpful in family counseling, since they signal patients’ acceptance of their mortality and represent a permission slip for the family to let go when the time comes. Their value is self-evident in New York state, where doctors are prohibited from removing life support from incapacitated patients without official proxies or clearly documented wishes. Such patients can end up stuck on breathing machines in a cruel and costly medicolegal limbo. Tragically, mandated life support of terminal patients inevitably leads to painful complications like skin breakdown, catheter-associated infections, and ventilator-associated pneumonia, for which Medicare no longer wishes to pay. In Texas, families who insist on care deemed “medically futile” can be overruled, but this
is controversial\textsuperscript{10} and no other state has granted physicians comparable authority over end-of-life decisions.

### Supply-Sensitive Care

Wennberg and coworkers\textsuperscript{4,14} focus on “supply-sensitive” care because their research has consistently demonstrated that geographic variations in spending parallel the local supply of physicians, medical specialists, and hospital beds, but appear unrelated to the health status of the population. They estimate that bed supply “explains” more than half of the variation in hospital admission rates for medical conditions,\textsuperscript{14} and they postulate that “any expansion of capacity will result in a subtle shift in clinical judgment toward greater intensity.”\textsuperscript{12}

To clarify whether care is driven more by illness or by physician and bed supply, Fisher et al\textsuperscript{3,4} studied resource utilization for 3 acute illnesses using a “natural randomization” technique to control for disease severity. Patients hospitalized nationwide with myocardial infarction, hip fracture, and colon cancer were retrospectively grouped according to which of 306 hospital referral regions they happen to reside in. When divided by quintiles of regional 6-month end-of-life spending, these national cohorts displayed similar baseline case mix, yet patients from the highest spending regions received \(\approx 60\%\) more care than those from the lowest spending regions, while both quality scores and outcomes were slightly worse. The authors reasonably theorize that more aggressive hospital care can worsen outcomes by exposing patients to greater risks of procedural complications, hospital-acquired infections, and medical errors.

This elegant study demonstrated that in the aggregate, the highest levels of spending include substantial amounts of wasteful and possibly harmful overtreatment. However, the association of high end-of-life spending with overtreatment among large cohorts of comparable illness severity does not prove that end-of-life spending itself is independent of illness severity, as is assumed in the Atlas. Fisher’s results also do not exclude geographic variations in other reasons for care or ensure that any category of care will be evenly distributed after further subdividing the population from national and regional levels down to individual hospitals, which vary greatly in local conditions, clinical services, and referral patterns. Such extensive subgrouping commonly introduces bias, which cannot be uncovered without clinical details.\textsuperscript{16} As Virnig\textsuperscript{31} has written, “a fundamental challenge faced by studies of intensity of end-of-life care . . . is that the observation of geographic variation does not provide evidence of its cause.”

### Comparing Hospitals

The Atlas data are now available on Consumer Reports’ new Web tool, which lists 2857 non–Veterans Administration hospitals by region (without reference to their type or size), rates them on a scale from 0\% to 100\% (from the most “conservative” to the most “aggressive”) based on the average number of inpatient days and physician visits over the last 2 years of life, and also compares out-of-pocket costs.\textsuperscript{11} The website warns consumers about the dangers of overaggressive care, lists commonly overused tests and treatments, and counsels patients to be more assertive in discussing the benefits, risks, and alternatives of procedures with their physicians. This is excellent advice, but the website’s presentation of the Atlas data still raises concerns.

The website discloses few limitations, except to say that emergencies like appendicitis will be treated appropriately (ie, aggressively) at every hospital. It incorrectly claims the data were “statistically adjusted . . . to account for how sick patients were,” whereas the authors only corrected for the type of chronic condition, plus basic demographics. Consumer Reports does mention one source of error: for patients treated at multiple locations, the Atlas assigned all data to the hospital visited last or most often. The website also acknowledges that its tool is only one way of comparing hospitals and is not designed as a quality indicator, and it provides a link to the government’s hospital quality data at www.hospitalcompare.hhs.gov, which are not concurrent.

The Web tool ranks all hospitals on a single scale, which cannot fairly compare institutions that provide different clinical services to dissimilar populations. As Wennberg et al\textsuperscript{4,14} have written, “typically, hospital level comparisons are confounded by differences in case mix across communities.” For example, referral bias is known to confound comparisons of costs and outcomes at tertiary centers with those at community hospitals.\textsuperscript{16} Patients who want conservative care may elect to be treated at and discharged from their community hospitals, whereas patients who are candidates for aggressive treatment such as cardiac surgery—especially complex and high-risk cases—may be selectively referred or self-referred to tertiary centers. Consequently, community hospitals will earn lower aggressiveness scores, because the proportion of deaths occurring after aggressive versus conservative hospital care will be lower than at tertiary centers.

This issue is highlighted in Table 1, which compares the data for New York University and University of California, Los Angeles with those for the Mayo Hospitals and the remaining 4 hospitals in the Rochester, MN, region. For several measures, differences between the 2 Mayo facilities (Rochester Methodist and St Mary’s) and the other hospitals within its region (most of which are in the Mayo Health System and presumably refer their sickest and most complex cases to the mother ship) are almost as great as differences between Mayo and the other academic centers. Because Mayo’s well-known efficiency is considered a benchmark, it seems likely that the more aggressive care at Mayo hospitals, compared with their neighbors, reflects disparities in illness severity and patient preference, and it follows that comparisons among academic or community hospitals may be similarly confounded by geographic variations in unmeasured variables.

If Consumer Reports’ rankings are adopted by payers, hospitals wishing to improve their scores might find new ways to boost efficiency. However, the ratings also could discourage providers from treating high-risk patients and might encourage more referrals to tertiary centers. This is exactly what was reported when mortality scorecarding for cardiac surgery was first introduced without proper risk adjustment.\textsuperscript{32} Yet, such referral patterns would only exaggerate the kind of variation that the Atlas considers unwarranted.
AHA/ACC Standards for Efficiency Measures

Because “enthusiasm for measuring and improving efficiency is not matched by a consensus regarding (methodology),” the AHA and ACC recently established criteria for assessing the suitability of efficiency measures for public reporting. These standards group the preferred attributes of efficiency measures into 4 domains including (1) integration of quality and cost, (2) valid cost measurement and analysis, (3) minimal incentive to provide poor-quality care, and (4) proper attribution of the measure.

According to the standards, commonly reported measures of resource allocation have important deficiencies (Table 2). Episode-based indicators including length of stay (LOS) and hospitalization costs can be risk adjusted but fail to relate the cost of care to its quality or outcome, thus poorly describe its value. Combination of a cost measure like LOS with an outcome measure like readmission rate (concurrently in the same patients) better expresses the quality and value of care and helps to ensure that reductions in LOS are achieved through more efficient care rather than by discharging patients prematurely. Risk adjustment remains necessary, since sicker patients may require longer LOS and more readmissions. Furthermore, valid cost analysis depends on the quality of data used to adjust for illness severity and comorbidities and on having enough observations per provider, which complicates evaluation of individual practitioners, as opposed to larger health systems. Yet, private insurers still use crude billing information for punitive cost profiling of individual physicians, who may be perversely tempted to “deselect” difficult patients before being deselected themselves.

The highest rated measure was avoidance of nonrecommended procedures, as defined by appropriateness criteria for when and how often to use tests and treatments for specific indications, based on scientific evidence and expert opinion, which are being promoted by the ACC as a means to improve our “stewardship” of Medicare dollars. Enough patient characteristics are intrinsic to the measures that the need for post hoc adjustment is reduced. For example, ACC working groups have identified as inappropriate 13 of 52 potential uses of myocardial perfusion imaging and 23 of 72 uses of cardiac CT and MRI. Based on these efforts, a pilot project involving professionally endorsed criteria for cardiac imaging was included in the recently enacted Medicare Improvements Act of 2008. This program offers providers decision support, feedback, and incentives for good compliance. Recently, the government suggested more aggressive cost-control practices like preauthorization, but the ACC anticipates that

<table>
<thead>
<tr>
<th>Measures</th>
<th>Quality-Cost Integration</th>
<th>Valid Measurement and Analysis</th>
<th>Minimal Incentive for Poor-Quality Care</th>
<th>Proper Attribution of the Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of stay</td>
<td>No*</td>
<td>N/A</td>
<td>No†</td>
<td>Yes‡</td>
</tr>
<tr>
<td>30-Day readmission</td>
<td>Yes and No§</td>
<td>N/A</td>
<td>Yes and No¶</td>
<td>Yes¶</td>
</tr>
<tr>
<td>Hospitalization costs</td>
<td>No</td>
<td></td>
<td>Yes</td>
<td>Yes‡</td>
</tr>
<tr>
<td>Nonrecommended tests</td>
<td>Yes††</td>
<td>N/A</td>
<td>Yes§§</td>
<td>Yes¶</td>
</tr>
</tbody>
</table>

*Length of stay is a measure of utilization with only an indirect association with quality.
†Incentive to lower length of stay could lead to premature discharge and adverse events including higher overall costs.
‡Attribution to the hospital is appropriate.
§Readmission indirectly incorporates considerations of cost and quality; however, cost of initial care is not included, and if extra resources were required to reduce readmissions, a singular focus on readmission would miss it.
¶Incentive to reduce readmissions could lead to behaviors that reduce access to the hospital for those who were recently discharged.
††Attribution to the hospital is appropriate, although there are also outpatient factors that are important.
†‡A singular focus on cost does not include consideration of quality.
§§Depends on methodology.
¶¶A focus on cost may lead to incentives to reduce necessary services and increase risk for adverse consequences for patients.
†††Unnecessary tests are costly and represent poor-quality care.
§§§Incentive is to avoid unnecessary testing.

“peer feedback reports will have a tremendous impact on inappropriate use,” and believes that preauthorization is not an “effective educational tool” and should only be considered for physicians who remain noncompliant despite education and feedback (R. Brindis, personal communication, 2008).

Appropriateness criteria may need encouragement in specialties where practice guidelines are not as well established. Among hundreds of multispecialty quality measures endorsed by the National Quality Forum41 and the American Medical Association,42 very few relate to avoidance of nonrecommended care. Furthermore, inappropriate procedures may only represent the tip of the iceberg of unnecessary care. Wennberg et al12 argue that most overtreatment occurs in the “gray areas” of medicine that are left up to clinical judgment (and, I would add, patient preference). This category includes care of “uncertain” appropriateness, or having an AHA/ACC class 2 indication, or outside of current guidelines—such as costs to the hospital visited last or most often, and partly because it ignores the role of patient/family preference and other provider-independent factors. In other words, I suggest that proper cost attribution should acknowledge the many social, cultural, and legal factors that drive care, especially near the end of life.

### End-of-Life Aggressiveness Rated According to the Standards

By the AHA/ACC standards, the end-of-life aggressiveness scale rates poorly (Table 3). For reasons discussed earlier, the intensity or cost of care in the last 2 years of life rates no better than LOS or episode-based hospitalization cost on quality-cost integration (domain 1) or poor-quality incentives (domain 3), and is inferior to all rated measures in the other domains.

With respect to valid cost analysis (domain 2), the measure rates negatively because, unlike episode-based indicators, looking back 2 years before death does not identify a clinically definable population. Of note, the Atlas method does not distinguish treatable stages of illness from terminal illness. Also, risk adjustment is impossible without an appropriate reference time to derive covariates and then track outcome; and the need to wait for a sufficient number of new deaths limits timely reporting of performance improvement.

### Looking Back at Decedents Versus Forward With Cohorts

Cost control is an enormous challenge requiring new methods to define and deter unnecessary and marginally effective care. I commend Wennberg et al12 for their forceful activism and agree that many “chronically ill and dying patients are receiving too much care, more than they . . . actually want or even benefit from.” I also support their prescriptions for strengthening primary care, for improving care coordination, for educating patients and physicians that more care is not always better, and for more research.

However, as currently defined, the Atlas model of relative provider aggressiveness has serious limitations. Numerous confounders make the cost or intensity of end-of-life care a very indirect measure of provider performance which, like other crude measures of resource allocation, will unfairly penalize those who treat sicker patients, offer more advanced treatments, or practice in social, cultural, or legal environments that foster higher levels of care.

Some authors suggest that retrospective studies of end-of-life care could be improved by shortening the look-back interval to 3 months or less, focusing on clearly terminal
cases (ie, metastatic cancer), and excluding unexpected deaths, or else by tracking cohorts instead of decedents.18,43 Decedents may be suitable for evaluating the process of care, but cohorts appear preferable for measuring efficiency, and would harken back to the classic Wennberg studies that identified unwarranted variations in the use of specific procedures such as tonsillectomy and hysterectomy.6

The Medicare Group Practice Demonstration is an ongoing cohort-based initiative designed to measure and incentivize both efficiency and quality.44 This project enrolled 10 large group practices, defined efficiency as a reduction in spending growth compared to a risk-matched community control group, and offered participants a share of their savings as well as quality bonuses. After 2 years, the groups saved Medicare an estimated $34 million on care of beneficiaries with diabetes, coronary disease, and CHF while maintaining excellent scores on quality measures such as diabetes control.45 The payment formulas need revision since, for most groups, administrative costs exceeded their bonus payments, but the concept that cohort-based, risk-adjusted cost measures and incentives can deter unnecessary care appears valid.

Conclusions

We have reviewed evolving provider efficiency measures and incentives. As per AHA/ACC guidelines, it is not sufficient to simply reward savings and hope that quality and outcomes are maintained. When possible, unnecessary and inefficient care should be defined prospectively and avoided systematically. These complementary approaches are being tested in current Medicare demonstration projects, as discussed earlier (Table 4). Compared with risk-adjusted cost measures, appropriate use criteria are more specific, much easier to implement (as evidenced by reports of voluntary self-monitoring),46,47 and more suitable for evaluation of individual practitioners, but more limited in scope. To promote more efficient care, redesign of clinical protocols and reallocation of resources also may be required.48

Near the end of life, clinical priorities change and measures of care should be adjusted accordingly. When illness becomes irreversible, clinical deterioration may no longer justify escalation of care, so risk-adjusted cost analysis may become less relevant than better definition and proper attribution of overaggressive end-of-life care. To avoid desperate overtreatment, better implementation of palliative care is needed, and hospital investment in such services appears fruitful.49 These efforts require measures that are clinically meaningful and actionable, such as documentation of advance care planning, availability of critical care consultation prior to intensive care unit admission, and access to ethics committees for dispute mediation.50 Performance improvement also depends on public backing, because end-of-life care that is more realistic and compassionate still lacks support on various cultural and legal fronts.

Acknowledgments

The author thanks Muriel Gillick, MD, Alan Moskowitz, MD, Kenneth Prager, MD, and Thomas Rosenthal, MD, for their critical review of the manuscript.

Disclosures

None.

References

22. Belluck P. Even as doctors say enough, families fight to prolong life. NY Times Health Section, March 27, 2005.


---

Key Words: appropriateness criteria • cost analysis • efficiency measures • end-of-life care • hospital costs • risk-adjustment
The Cost of End-of-Life Care: A New Efficiency Measure Falls Short of AHA/ACC Standards
Gerald W. Neuberg

doi: 10.1161/CIRCOUTCOMES.108.829960
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/2/2/127

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/