Alvan R. Feinstein, M.D.

Appraising the quality of medical care has been a dynamically growing activity for the past three decades. Although efforts to evaluate and improve care were made by Florence Nightingale (Nightingale, 1863) in the nineteenth century and by Ernest A. Codman (Codman, 1917 [1996]) early in the twentieth, the recent growth in the United States has been spurred by financial incentives. The federal government, commercial organizations (such as insurance companies), and private foundations have all sought to reduce the increasing costs of care while hoping to maintain satisfactory quality.

The incentives offered by these groups have spawned a large array of organizations and programs that have struggled with the challenges of deciding what to regard as quality of care, choosing ways to measure it, and developing improved mechanisms to deliver it. The challenges have been complex and difficult because quality is like both beauty and pornography: like beauty, it is often in the eye of the beholder; like pornography, it is often hard to define but easy to recognize.

My goal in this chapter is to indicate that the total activities of care contain elements of curing and caring; but almost all existing procedures for appraising “quality of care” today are aimed only at curing. Because the existing procedures cannot adequately evaluate caring, a new approach is proposed at the end of the chapter.

Diversion of Goals and Concepts

Efforts to define quality of care have frequently used the “avoid-and-divert” policy that is common in modern medical science. For example,
the search for a suitable definition of excellent clinicians has often been diverted to the demand that they pass an appropriate certifying examination. Analogously, the question of what to regard as quality of care has usually been avoided by diverting the response to answers available from diverse measurements and mechanisms of health care delivery. Thus, the easily measured rates of mortality or postoperative complications have often been used to indicate the quality of surgical procedures; and the easily ascertained existence of appropriate facilities and numbers of personnel has denoted the quality of a hospital or other health care system.

This avoid-and-divert strategy is not new or unusual. For most of the twentieth century, as individuals, groups, and societies have been both benefited and harmed by the impact of magnificent technology, investigators in the social, behavioral, and clinical sciences have generally preferred an exact answer to a diverted question rather than an imprecise or approximate answer to the direct question. More than fifty years ago, the strategy was summarized for the social sciences in Frank Knight’s memorable aphorism: “If you cannot measure, measure anyhow” (Wirth, 1940: 169).

Perhaps the most obvious diversion has been the altered concept of the word care itself. The traditional medical idea is that care denotes concern, empathy, compassion, and other Samaritan aspects of the doctor’s attitude toward the patient (McDermott and Rogers, 1983). Today, however, particularly as doctors have come to be called providers, care usually refers to the array of activities performed by providers in “managing” the patients, who are often called consumers. The interpersonal activities once called care are now usually called caring, and the phenomena subsumed as care generally refer to entities that might best be called curing.

The change in terms from doctors to providers is not unreasonable, since many persons other than physicians regularly contribute to the activities of care. The use of care to refer to management rather than caring is also not unreasonable, since care is probably the best simple word for denoting the total scope of activities in preventing or treating disease.

The word disease has also changed, of course. Before the nineteenth century, it referred to a patient’s discomforted ease—that is, dis-ease—but it now denotes a more abstract entity, often manifested by morphologic or laboratory abnormalities that were not discernible before the advent of modern technology. (The symptoms and other direct clinical manifestations that were once called dis-ease are now usually cited as illness.) Even the change from patient to consumer is also defensible since patient comes from the Latin, pati, which means “to suffer”; and not all persons receiving care...
are in a state of suffering. Nevertheless, the many changes in nomenclature and usage reflect the profound alterations of a long-standing medical tradition. A suffering patient who received caring has been replaced by a consumer whose diseases are to be cured by preventive or remedial management. The activities of that management are what become appraised today as quality of care. The appraisal procedures might be criticized for not reflecting the goal of caring, but they are a product (and reflection) of major changes that had occurred in the world of medical “care” long before the appraisal procedures were developed.

**DOMAINS OF QUALITY OF CARE**

To facilitate measurement, quality of care has been divided into different domains that might be cataloged as components, constituents, and standard setting, and that are approached with different methods for expressing and evaluating the contents of the domains. An early and now often-used approach was proposed when Donabedian (1980, 1982) divided the domains of quality of care into three main components. The components could then be subdivided into diverse constituents; and standards could be set for evaluating the constituents.

**Components**

Donabedian’s three main components were called structure, process, and outcome. **Structure** refers to the personnel, facilities, and system involved in health care delivery. **Process** refers to the actions that emerge from those structural elements. **Outcome** denotes the subsequent events that follow the process. For example, a nurse practitioner (as part of structure) might recommend (as part of process) an analgesic for the treatment of headache. As outcome, the headache might or might not be relieved.

This classification produced an excellent basic taxonomy for demarcating the entities to be considered, but an immediate problem was apparent. The three sets of entities occurred sequentially, but were not consequential. Effects could not always be attributed to or associated with the antecedent “causes.” Thus, when a factory manufactures inanimate objects, an excellent structure and process should always be followed by an excellent outcome. In the vicissitudes of the medical ailments and personal interchanges of human life, however, a good structure may be accompanied by a poor process, and vice versa. Furthermore, a good outcome may occur despite inadequacies in both structure and process; and a
splendid structure and process may sometimes be followed by a poor outcome. For example, my great-grandmother delivered nine healthy babies at home (or in a farm field) without any special prenatal care or medical attendance. Conversely, a stillbirth can occasionally occur today despite excellent “care” during the pregnancy, accompanied by multiple sonograms, superb hospital facilities, and a board-certified obstetrician.

Because of the dissociations and uncertainties, an unresolved conflict has existed about where to focus the main evaluation of quality of care. On one side are those who argue that only outcomes are important; on the other side is the argument that only structure and process can be really improved because outcomes are too dependent on unalterable (and often unpredictable) biologic phenomena of disease and human beings. What are often ignored during the argument, however, are the different constituents of structure, process, and outcome, and the interrelationships among those constituents.

**Constituents**

Each of the three component domains contains constituent entities that have both form and function. The forms can usually be easily identified, specified, and documented. Thus, in a particular health care system, we can count such structural elements as the number of registered nurses, board-certified doctors, and transportation orderlies, and the existence of a hospital ombudsman for patients’ complaints. For a specified clinical condition, we can document whether an inappropriate treatment was used, or whether an appropriate treatment was omitted or given at the right time. To address the form of an outcome, we can identify whether the patient survived, whether blood pressure or blood lipids were reduced, and whether a low hematocrit was raised.

The form of each of these constituent entities, however, is accompanied by behavioral and functional phenomena that are more difficult to measure. Are the nurses and doctors empathic and compassionate? Do the transportation orderlies take patients to a radiologic suite and then abandon them in the hallway? Is the ombudsman merely a titular position or is it someone who is truly available, empowered to act, and able to evoke appropriate remedial actions?

When pursuing the form of the orders placed for appropriate (or inappropriate) diagnostic tests and treatments, do the hospital personnel suitably explain their decisions and prepare the patients for what is going to happen? Is the form of the outcome entities restricted to survival and eas-
ily obtained technologic measurements, or do they include such functions as the patient’s anxiety, discomfort, physical capacity, and satisfaction with care?

These behavioral and functional phenomena are much more difficult to measure than the existence and dimensions of form for constituent elements. Consequently, the functional phenomena are seldom measured, and, when “measured,” are seldom checked for personal interrelationships. Accordingly, the close correlation that might occur between certain processes and outcomes is seldom documented; and the discussion may be relegated to the scientifically disreputable status of anecdotes. For example, although death may be counted as a poor outcome despite appropriate therapy, the patient’s serenity and the family’s tranquility during the period of dying may be an excellent outcome that can be directly associated with thoughtful, compassionate care by the medical and nursing staff. The anecdotes may also include many instances in which a disgruntled patient and family took legal action against an arrogant surgeon, despite an apparently good outcome, and yet other instances in which a patient and family, happy about sensitive attention from the health care system and personnel, did not seek legal redress for gross errors in therapy.

The form/function distinction in the constituent elements also reflects the difference between caring and curing. The elements of form in the existence or dimensions of specified constituents are usually measured in reference to such curing outcomes as survival and restoration of physiologic normality. The functional performance of the elements, however, is seldom examined for their contributions to such caring outcomes as serenity, tranquility, and good quality of life. Consequently, the possibility that caring outcomes might be highly correlated with structure and process is usually missed when the correlation measurements are restricted solely to outcomes aimed at curing.

Standards of Evaluation

Setting standards of evaluation is often regarded as the most difficult part of appraising quality of care, not just because the necessary information may be omitted from collection, but particularly because of disagreements about how the standards should be set and applied. The topics that give rise to these disagreements can be divided into two main parts: ideologic disputes and methodologic problems in statistical expressions. The topics are extensive enough to warrant separate discussion in the next few sections.
IDEOLOGIC DISPUTES

The disputes labeled here as ideologic arise from ideas or doctrines that cannot be documented as proved. They represent beliefs and preferences, such as choice of vocation, avocation, religion, political party, or type of entertainment. Each set of preferences can be accompanied by its own justifications and rationales, and each choice may be worthwhile or self-defeating, defensible or attackable. The choices themselves, however, cannot be demonstrated as either right or wrong, although the decisions can profoundly affect the way a person lives, enjoys life, and dies.

In the field of quality of care, the ideologic decisions arise from the choice of viewpoints, criteria, scope, and grading of the evaluations.

Viewpoints

Whose viewpoint should be used when quality of care is evaluated? For sociologists and economists, the view is usually aimed at society or systems of health care. This view is encouraged by national organizations that are concerned with public expenditures, corporate profits, or such general goals as universal access to care.

The clinical and public health investigators who examine quality of care, however, usually work at (or get data from) a particular institution. They often know the associated health care personnel and may enlist them in consultation or collaboration. The research activities may then use the viewpoint of the professional personnel.

A third viewpoint is that of the individual persons who receive the care. This viewpoint might seem appropriate and desirable, but is seldom used. The main argument against it is that most patients are regarded as not qualified to make decisions about the efficiency of a system, the competence of the personnel, the appropriateness of therapy, or the appraisal of a clinical outcome. Besides, many patients may be too subjective, inconsistent, or nonrational to make delicate judgments. (Very little evidence exists, however, to demonstrate that patients are necessarily more subjective, inconsistent, and nonrational than the concomitant health care personnel.)

The omission of the patient’s viewpoint may seem to make things more scientific, but it also tends to make curing rather than caring the main focus of evaluation. The system and the professional staff can readily use their background knowledge and foreground goals to examine curative phenomena; but caring is usually best discerned by its recipients.

An additional problem is that many acts of caring are difficult or com-
plex to describe; and they are seldom entered in the medical records from which care is often evaluated. An arm across the shoulder, a hug, a squeeze of the hand, a subtle perception, a compassionate response, a useful explanation, a thoughtful comment to the patients’ spouse or child, a rearrangement of the bedclothes, an appropriate drawing of curtains or lowering of voice—these can all be important acts of caring. Nevertheless, unlike the “curative” data of laboratory and technologic procedures, the information about these elements of caring is seldom entered in the medical record. Having devoted the extra time to do the caring, its provider may not want to take additional time to record the details. Furthermore, in some of the new-style computerized medical records, which allegedly save time (i.e., allow more time for acts of caring), no format may be available to report them, even if the provider wants to do so.

Criteria

Separate topics of dispute are the approach used for setting criteria, and the choice of the standards to be applied in grading.

To evaluate care for a particular patient, we would need to consider everything that happened: the presenting manifestations, the first set of clinical decisions (often a diagnostic work-up), the next set of clinical decisions (often a choice of consultation or therapy), and all the subsequent decisions, actions, and events. Because each patient’s situation will usually have its own individual distinctions, standard criteria could not be developed for routine application to all possible cases. The appraisal of each patient’s care could still be done, however, if suitable expert appraisers review what happened and apply their own implicit criteria and standards during the evaluation. (This type of evaluation also occurs during the oral examination of a candidate for board certification or for higher academic degrees.)

The use of implicit criteria has the advantage of being aimed at the right thing—the total care of a patient—but it is associated with many scientific disadvantages. The appraisers who do the work may not always be suitable in skill, temperament, objectivity, and consistency. In the absence of specifications for the criteria, the different appraisers may reach different decisions for the same case; and whatever is learned during the evaluation process cannot be identified and cited for justification or for application elsewhere.

To obtain the detailed specifications needed for explicit criteria, however, the scope of evaluation must be sharply narrowed. Instead of check-
ing an unselected series of patients, the appraisers may check a highly se-
lected set of tracer, or indicator, conditions (Kessler, Kalk, and Singer,
1973) that are reasonably common and important in a large group of pa-
tients. In primary-care practice, for example, these conditions might in-
clude sore throat, urinary tract symptoms, back pain, and hypertension.
In a consultant or hospital practice, the conditions might include acute
myocardial infarction and acute gastrointestinal bleeding. After the tracer
conditions have been selected, explicit criteria can be constructed and ap-
plied to appraise the management of each condition. (In a modern exten-
sion of the activity, such criteria are now prepared and offered not just for
retrospective evaluation of care, but particularly for prospective instruc-
tions to the providers. The new criteria are called guidelines, or critical
pathways.)

Scope

Explicit criteria have the obvious scientific virtues of being stipulated, up
front, and applicable by diverse evaluators, not just by experts. On the
other hand, the criteria pertain only to the selected conditions and are of-
ten applied, despite all the detailed explications, with substantial inter-
rater variations (Goldman, 1992). Perhaps the greatest disadvantage, how-
ever, is in scope. The specifications in the criteria are almost always aimed
at acts of testing, ordering, and performance that can easily be determined
in the medical record. Consequently, the explicit criteria become yet an-
other mechanism for emphasizing activities of curing rather than caring.

This same disadvantage, of course, occurs with implicit criteria, since the
main source of the difficulty is the use of medical records, which seldom
contain the information needed for the evaluations. Nevertheless, an expert
who appraises the patient’s entire record, rather than just the segment
needed for management of the tracer condition, may sometimes find acts
of caring mentioned in nurses’ notes, clinical progress notes, or other loca-
tions (for many medical records, the expert appraiser must first be able
to master the formidable challenge of reading handwriting before making
any additional discernments). In such sites as office practices, nursing
homes, or hospices, however, the medical records often have a much more
limited set of entries; and for patients treated at home, no record may be
available.

For all these reasons, caring (rather than curing) cannot be evaluated
from information recorded by the providers of care. The evaluation proce-
dure for caring must either contain direct observation and monitoring or
must rely on retrospective inquiry from recipients. Direct observation is expensive and time-intensive; and retrospective inquiry and subjective reports will raise hackles in the scientific community.

**Standards and Grading**

The fourth set of ideologic issues refers to the setting of standards and grading of the evaluations.

Before the standards are set for evaluating care, decisions must be made about what to evaluate. Should each case be checked in an individual, unselected manner or should there be a focus on patients with “tracer conditions”? If the latter, which conditions should be chosen? For either unselected cases or tracer conditions, should the entire course of care be reviewed or should it be only individual episodes? For example, if a patient has recurrent episodes of congestive heart failure or asthma attacks, the care given between episodes may sometimes be more important than what is used for each episode. For a patient with persistent chronic conditions such as hemiplegia, urinary incontinence, or dementia, the criteria and evaluation procedure will be quite different from what is pertinent for asymptomatic hypertension, diabetes mellitus, or hyperlipidemia. Furthermore, although most of the efforts devoted to setting standards have been aimed at curative topics, issues in caring are particularly pertinent for patients with permanent disabilities or with ongoing impairment in activities of daily living.

Assuming that these decisions about what to evaluate have been satisfactory, we can then proceed to the standards used for the evaluation. When acts of care are deemed appropriate or inappropriate, excellent or poor, laudable or reprehensible, who should choose the grades used for the ratings and the standards that elicit a rating for each grade?

Should the standard setters be expert consultants? Ordinary practitioners? Nurses? Economists? Or the members of the public who receive the care? Elitist standards set by expert consultants at tertiary care institutions may be appropriate for those institutions but unsuitable for the work of representative practitioners elsewhere. For example, during my six-year term as a member of the governing group of the American Board of Internal Medicine, I not once succeeded in getting the written examination to include a question for which the answer was “do nothing” or “watchful waiting.” Although proposed as a response to several clinical scenarios, this answer was unacceptable to many of my specialist colleagues in the committee. In each scenario, they regularly found something they would do.
An effective way of choosing the standard setters might be to focus on the entity being evaluated. If it represents a curative action, the standard might be set by persons most appropriate to evaluate the merits of the action and its pertinence in different clinical situations and locations. If the action is related to cost-efficiency ratios, the decisions might be made by economists or a clinicoeconomic collaboration. If the entity is a caring phenomenon, however, the standards might be best set by patients and family, perhaps aided by skilled nurses.

Regardless of who sets the standards, their role requires separate consideration. Are they intended to give general grades to all acts of care and perhaps to elevate the average performance of care? Or are they supposed to detect the outlier providers who are particularly poor performers (an activity sometimes called “getting the dangerous drivers off the road”)? If the goal is to find and perhaps improve or remove the undesirable outliers, the standards might be set in a simple pass/fail manner according to egregiously unacceptable performance of certain critical challenges.

For example, analgesics for an ordinary tension headache might be regarded as an appropriate recommendation in ordinary primary care, but some other mechanisms of relief (such as rest or whiskey) might not be greatly downgraded or deemed grossly inappropriate. Consultation with a neurologist, and particularly with a neurosurgeon, however, might be regarded as an outlying, improper decision. On the other hand, suppose the patient has had recurrent tension headaches, is emotionally alarmed about them, and insists on more elaborate care? Is the doctor then remiss in ordering a brain scan or an expert consultation?

Finally, the last challenge to be cited in setting standards is the external problem of whether sins of commission are more serious than those of omission. In the old, pretechnology medical era, the guiding principle was *primum non nocere* (above all, do no harm). Today, however, with so many powerful but sometimes harmful agents available, the doctor who always seeks to avoid harm will also fail to provide important benefits. A more frequent guiding principle today, therefore, is to seek an optimum risk/benefit ratio. Nevertheless, what may be deemed an appropriate omission may sometimes lead to adverse effects, and sins of commission may sometimes have major benefits. For example, giving antibiotics for the common cold is usually condemned as grossly inappropriate. On the other hand, in patients who apparently have a common cold, many Group A streptococcal infections can exist without producing the classical manifestations of pharyngitis and can be successfully treated by the an-
tibiotic. In fact, the virtual disappearance of acute rheumatic fever in the United States is often ascribed to the overuse of antibiotics that were given for the common cold but that, incidentally, eradicated many potentially noxious streptococci. The practitioner who does not give the antibiotic may be lauded, but the acclaim may not be shared by the rare patient whose untreated streptococcal infection leads to rheumatic fever or glomerulonephritis.

On the other hand, a doctor may get full credit for excellent practice in the sphere of curing if a patient with acute myocardial infarction is admitted to an intensive care unit, monitored with all the pertinent equipment, and given all the appropriate medication while in the hospital and at discharge. The doctor may not, however, have offered any words of caring assurance while the patient was terrified by associated events in the intensive care unit, and may not do so later when the patient is at home after discharge, an invalid, fearing death at any moment.

STATISTICAL METHODS

Another reason for diverting quality-of-care assessments to topics of curing rather than caring is that curing can much more easily be cited in terms currently amenable to quantitative statistical expressions. The use of statistics in medicine, rudimentary when the twentieth century began, has now become a customary and often dominant activity. The quality of a nation’s health care is often rated according to statistical data for longevity and infant mortality rates. A hospital’s quality of care may be cited according to mortality rates and counts of personnel and equipment. New drugs will be licensed only if accompanied by “statistically significant” data from randomized clinical trials; and meta-analytic statistical aggregations of randomized-trial data have become the basis for what is now called evidence-based medicine. The considerable benefits produced by the statistical activities are well recognized and beyond dispute, but like all other potent interventions, they have had many adverse side effects.

Humanists often complain that modern medicine has become too oriented toward reductionist science. To explain mechanisms of disease and human biology, the scientific methods have progressively reduced the investigated entities downward from intact organisms to organs and systems, then to tissues and fluids, and ultimately to membranes, cells, and molecules. Since the study of intact people is essentially eliminated in this approach, reductionist medical science is often accused of being dehumanized.
On the other hand, the reductionist approach is entirely appropriate for the goal of exploring mechanisms of biologic action, and the approach has led to magnificent scientific advances in technology, in therapy, and in understanding diverse mechanisms of development for human ailments. What is often overlooked during the attack on biologic reductionism is the much greater antihumanistic threat of statistical reductionism by researchers who are apparently studying patient care.

Statistical reductionism—as discussed in the next few sections—occurs with mathematical or psychometric methods that (1) rely on “hard” data, avoiding the crucial “soft” data of distress and caring; (2) express complex phenomena not as a coordinated whole but as an aggregate of multiple items; (3) purposely ask questions in a general, depersonalized manner; (4) deliberately eliminate important personal details; and (5) report results for an “average” person.

**Reliance on Hard Data**

Most studies of therapy have relied on outcome events delineated by so-called hard data, such as death, diagnoses of disease, and laboratory or morphologic data. The statisticians are justified in wanting to use “reliable” data and avoid the soft information of symptoms, patterns of illness, and other crucial personal phenomena. The focus on hard data, however, helps produce a dehumanized set of information and also allows clinical personnel to continue avoiding the challenge of making the soft data respectable. The respectability could easily be attained if clinicians acknowledged the importance of the challenge and then worked to improve the methods of observing and expressing the pertinent phenomena. A pioneer in this type of clinimetric approach almost fifty years ago was Virginia Apgar, who converted the implicit rating of a newborn baby’s condition into the explicit score by which she is commemorated (Apgar, 1953).

**Getting a Whole from Aggregated Parts**

The Apgar score converted a so-called soft rating for a whole phenomenon, such as excellent for a newborn baby’s condition, to a relatively hard aggregated sum of five parts (or items), each rated as 0, 1, or 2. The approach was successful because she chose the five parts according to sensible clinical judgment about importance; and the score she produced was easy to remember and to use.

The same strategy seems to have been applied in statistical and psychometric methods that also produce the rating for a whole from a sum of ag-
Aggregated parts. The results appear in multivariable analyses that predict complex phenomena, such as prognosis, by using such aggregated mathematical sums as $Y = b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + \ldots$, where the $X$ values represent individual variables, and the $b$ values are coefficients for the weights of each variable. The results also appear when other complex phenomena such as satisfaction with care, functional status, or quality of life are indexed as the sum of ratings for a series of multiple items that are regarded as constituents of the phenomena.

Although seemingly similar to Apgar’s basic strategy, the statistical and psychometric methods are carried out with striking clinical differences (Feinstein, 1999). Apgar chose a small number of crucial variables, using the “dissected intuition” of her clinical experience. The mathematical methods employ a much larger number of variables—usually thirty or more; their importance is determined not with clinical judgment but from mathematical calculations; and the variables often do not include crucial phenomena, such as clinical severity and comorbidity for prognosis, or spirituality and emotional satisfaction for quality of life. Apgar took a carefully considered whole and suitably demarcated its parts. The mathematical methods, after aggregating a set of arbitrarily chosen and arbitrarily weighted parts, claim that their sum is a rating for the whole.

Clinicians have traditionally used the Apgarian clinimetric approach when they asked the questions “How are you?” and then asked further questions such as “In what way?” or “What bothers you the most?” This approach has often been dismissed, however, as lacking the statistical credentials of reliability and validity. The substitute mathematical approach is conducted essentially as “Please enter your responses to this lengthy set of questions, and then I will tell you how you are.”

The conversion of a complex whole to a set of aggregated parts is the essence of reductionism; it is, however, often employed for such humanistic goals as determining satisfaction with care or quality of life. Since the procedure usually deprives patients of stating their own distinctive desires, expectations, and discontents, the failure to develop and apply a better technique is yet another manifestation of the avoid-and-divert phenomenon.

Depersonalized Questions

Another problem in the psychometric approach is the reluctance to ask direct questions (Feinstein, 1987). The direct approach might say, “Do you like your doctor? If not, why not?”; the psychometric tactic, on the
other hand, might have responses ranging from strongly agree to strongly disagree for statements such as “Most people like their doctors” and “Most doctors carefully explain and report results for the tests they have ordered.” The information may receive satisfactory statistical credentials, but it may not lend itself to separating out patients’ attitudes toward their personal doctors from their feelings about doctors in general; in this regard, interpretation may be impossible.

The Elimination of Important Personal Details

The essence of statistical analysis is the reduction and elimination of details. A simple example is the use of “three significant digits” to express the value of $\pi$; i.e., as 3.14, rather than 3.14159 . . . . A more compelling example is the reduction of a distribution of data to a single central index, such as the mean, or median. The reductionism can be highly valuable when applied to the decimal digits of a number or to offer the “univariate” summary value for a single variable, such as age, weight, or blood pressure. The reductionism is undesirable, however, when it eliminates variables rather than digits or distributions. A thoughtful caregiver—who will consider the patients’ age and weight and blood pressure and a great many other variables such as disability, distress, and social support—is not well served by results that not only eliminate the latter three variables but may not even include them in the collected data.

The Use of “Average” Results

The reduction of detail is a reasonable, necessary part of statistical analysis. No analyses could be done if all of a person’s individual details were always considered. Unless the details are reduced enough to allow suitable groups to be delineated, the groups will be too small for their membership to contain reasonably stable numbers that can be attributable to factors other than chance probabilities alone. Thus, if we want “statistically significant” results, we need big groups, but to get big groups, we must reduce the descriptive detail for each group.

This conflict produces another major dilemma in modern medical care. During the past half century, the use of randomized clinical trials has led to majestic advances in therapeutic science. In contrast to the anecdotes, authoritarian doctrines, and unquantified clinical judgments of the past, documentary evidence has become available, for the first time in medical history, to demonstrate the efficacy of therapeutic agents. The patients admitted to the trials, however, may all have the same basic clinical
condition or disease being treated, but may be otherwise heterogeneous in having diverse personal and pathophysiologic characteristics.

The randomization process may distribute those diverse characteristics in an equitable manner among the compared therapeutic groups, but the final results pertain to an average patient in each group. These average results have been splendid for the overall decisions made by regulatory agencies, pharmaceutical manufacturers, and other general evaluators of therapeutic efficacy. The results have also been valuable when efficacious agents were chosen for cost-benefit and cost-effectiveness analysis and when “guidelines” and “critical pathways” were established for general policy decisions in clinical practice.

The average results of randomized clinical trials, however, have not been satisfactory for clinical decisions about the care of individual patients. To demonstrate efficacy, a particular therapeutic agent may have been compared against placebo, not against other effective agents; and the trial data may not contain (or report) adequate information for the patients’ distinctive clinical and demographic subgroups. Like other major scientific and statistical advances, the average results may be excellent for public policy decisions, but not for individual decisions about patient care.

ACHIEVING MEASUREMENTS AND STANDARDS FOR CARING

For all of the foregoing reasons, the procedures that have been developed for measuring quality of care cannot be satisfactory for appraising success in caring. If such appraisals are desired, a new set of procedures will be needed. They can probably best be developed not by delegating the task to outside methodologists and other external consultants, but by persons who are intimately familiar with what actually occurs; that is, who know the needs and activities of caring. Those persons will be patients, caregivers, family members, and associated contributors, rather than more distant medical personnel.

A first step would be to ask the pertinent participants for particularly bad or good examples of caring. The questions should be asked in an open-ended manner, allowing a free range of topics and scope of comments in the responses. The examples that are mentioned should be reported not in general phrases such as “thoughtless,” “communicates well,” or “insensitive,” but in specific descriptions such as “abandoned me in the
hall outside the radiology suite,” “told me exactly what to expect when I went for the MRI,” or “was unaware of the religious and cultural anxiety of a devout Catholic admitted to a Jewish hospital.”

When enough specific examples have been accumulated, they can be organized into the more general categories of an appropriate taxonomy. Such a tactic was used to construct the taxonomic categories when Matthews and Feinstein (1988) prepared a “review of systems for the personal aspects of patient care” in a hospital, and when the taxonomy was later applied by patients to rate physicians’ performance in a hospital setting (Matthews and Feinstein, 1989).

The procedure may lead to a multi-item inventory that at first resembles the multiple items used in psychometric strategies. The inventory of caring, however, will have several major differences. The items will have been chosen, for both topics and importance, by the respondents, rather than by the investigators or by a mathematical rating system. The performance of the individual cited entities may be given simple binary ratings (yes/no or present/absent) rather than ordinal categories (such as strongly agree, strongly disagree) or ordinal frequencies (such as always, often, never). And the results can be used not to achieve an aggregated score of points but to denote individual problems that need solutions. In addition, the inventory itself can offer powerful details of phenomena to be considered when caring is taught or evaluated.

However, the current challenges in caring for patients with chronic illness, permanent disabilities, and premoribund conditions will require more extensive information than that obtained for hospital situations. The work will need input not just from patients but also from families, caregivers, and other pertinent persons. The standard constituents used for structure in Donabedian’s classification will also need expansion to include nursing homes, assisted-living facilities, adult day care centers, hospices, and other sites where caring occurs. The constituents should also include other participating personnel (nurses’ aides, visiting nurses, physical and occupational therapists, clergy, social workers, family members, close friends, volunteers) beyond doctors, nurses, and administrators.

After all the elements of caring have been identified, the criteria for “successful” caring should be established by recipients and the other participants; they are the people most appropriate for determining what is needed and wanted and for appraising how well the hopes and desires are fulfilled. The work of identifying the elements and establishing criteria can be greatly aided by assistance from a capable methodologist and a
knowledgeable clinician, who could help decide whether realistic goals are being set. The methodologist, however, should come with an open mind, prepared to discover reality, learn from it, and develop appropriate new ways to analyze it. The project can be greatly harmed rather than helped if the methodologist arrives with preconceived notions and rigid principles of conventional statistical or psychometric doctrines.

The challenges in such a project are new, having been inadequately approached by the methods of the past. Accordingly, the methods needed to master the new challenges will require new viewpoints and approaches. The same dimensional measurements used for a volume of urine, milligrams of sodium, or duration of survival will never be applicable for appraising acts of communication, compassion, or help. For people who believe these acts are worthwhile and important, however, they can easily be recognized, identified, classified, and organized into an effective taxonomy.

About a century ago, believing that chemistry was an artful but important entity, Mendeleyev constructed a fundamental taxonomy called the periodic table. His job was relatively easy because he needed to consider only two constituents: weight and valence. The result, however, was an essential basis for all the subsequent scientific progress that has now led to modern technology, with its magnificent achievements and sometimes lamentable consequences in humane medicine and in caring for patients.

Although caring is one of the oldest activities of human existence and medical practice, the procedures of caring have now become extraordinarily challenging, because they have been omitted not only from the education given to most modern physicians but also from the customary appraisals of “care.” Since the art of caring is well known to the people who give and receive it, the main challenge now is to identify, appraise, and teach the constituents of the art. It need not be made into a science per se, but an appropriate taxonomy of constituents can provide new dignity, respect, attention, and development.

REFERENCES


Codman, E. A. (1917 [1996]). *A Study in Hospital Efficiency*. Privately printed; reprint, Oakbrook, IL: Joint Commission on Accreditation of Healthcare Organizations.


