

Appropriateness in Patient Care: A New Conceptual Framework

VIRGINIA A. SHARPE
and ALAN I. FADEN

Georgetown University Medical Center

IN THE 1970S AND 1980S, DEBATES ABOUT HOW TO improve quality and contain costs in American health care focused on the identification and prevention of “unnecessary” care: specifically, unnecessary surgery. Although different definitions of what constituted an “unnecessary” treatment complicated investigation of the problem (Stroman 1979, 32), it was ultimately recognized that a major drawback of this focus was its implicit assumption that an intervention is effective until proven otherwise. In other words, it places the burden of proof on those who would contest an intervention, rather than those who would recommend it. As evidence increasingly showed that only a small minority of medical and surgical treatments had been subjected either to rigorous, controlled studies (Office of Technology Assessment 1983; Berwick 1989) or to effectiveness research (Brook 1991), the burden of proof began to shift. Recognizing that judgments about surgical overuse and the provision of unnecessary services depend on *prior* judgments about the criteria for *appropriate* use, the RAND Corporation inaugurated the field of “appropriateness research” in the 1980s with its Health Services Utilization Study.

The Milbank Quarterly, Vol. 74, No. 1, 1996
© 1996 Milbank Memorial Fund. Published by Blackwell Publishers,
238 Main Street, Cambridge, MA 02142, USA, and 108 Cowley Road,
Oxford OX4 1JF, UK.

In this article, we look at some of the challenges that accompany the conceptual and practical stress on the "appropriateness" of care. Although these challenges may extend over the range of health services, we have (following the historical attention to unnecessary surgery) limited our discussion to decision-making for surgical interventions. In particular, we focus on the evaluative nature of the concept of appropriateness and propose a framework for distinguishing between three sources of value that give meaning to this concept in patient care. The first is the *clinical point of view*. We argue that clinical recommendations demand a broader and more substantial evidentiary basis and that this information should be an integral part of the informed consent process. The second is the *point of view of the individual patient* in determining an intervention's "desirability." The third is the *societal point of view*, where appropriateness is understood in terms of the "cost-worthiness" of possible outcomes. This framework is also used to shed light on the contentious issue of medical futility.

The RAND Health Services Utilization Study

In the pioneering RAND studies, a procedure or intervention is designated "appropriate" when the "expected health benefit (i.e., increased life expectancy, relief of pain, reduction of anxiety, improved functional capacity) exceed[s] the expected negative consequences (i.e., mortality, morbidity, anxiety of anticipating the procedure, pain produced by the procedure, time lost from work) by a sufficiently wide margin that the procedure [is] worth doing" (Park et al. 1986).

On the basis of this definition and a detailed methodology for the development of recommendations by consensus, RAND researchers have, over the last decade, conducted numerous studies on the appropriateness of interventions like coronary angiography, gastrointestinal endoscopy, carotid endarterectomy, coronary artery bypass graft (CABG), percutaneous transluminal angioplasty (PTCA), and hysterectomy. On average, the performance of these procedures was rated appropriate in only 63 percent of cases (see table 1).

One of the purposes of the Health Services Utilization Study was to investigate the phenomenon of geographic variations in the use of surgical services. In a landmark study of 13 Vermont hospital service areas, for

example, Wennberg and Gittelsohn (1973) found that the number of procedures performed per 10,000 persons ranged from 13 to 151 for tonsillectomy, 10 to 32 for appendectomy, and 30 to 141 for dilation and curettage. RAND set out to test the generally held view that disproportionately high rates of use were indicative of unnecessary surgery. Their results, however, showed no such correlation (Chassin et al. 1987a). In fact, in the RAND studies, rates of inappropriateness ranged between 4 percent and 32 percent and were consistent over *both* high- and low-use areas (Chassin et al. 1987a). The most compelling explanation that has emerged for the phenomenon of geographic variation is professional uncertainty regarding the value of a procedure and the indications for its use (Wennberg, Barnes, and Zubkoff 1982; Eddy 1984). Widespread acceptance of this theory has produced a number of strategies to explicitly determine indications for the appropriate use of medical and surgical interventions. This is the mandate of the U.S. Agency for Health Care Policy and Research (Marwick 1993) and the catalyst for specialty society development of practice guidelines.

The RAND experience offers important insight into the challenges that accompany the conceptual and practical stress on the appropriate-

TABLE 1
RAND Appropriateness Studies

Year	Procedure	Number	Appropriate/ inappropriate (%)	Reference
1987	Coronary angiography	1677	74/17	Chassin et al. 1987b
1987	Gastrointestinal endoscopy	1585	72/17	Chassin et al. 1987a
1988	Carotid endarterectomy	1302	35/32	Chassin et al. 1987a
1988	CABG	386	56/14	Winslow et al. 1988
1990	Coronary angiography	320	49/21	Gray et al. 1990
1990	CABG	319	55/16	Gray et al. 1990
1993	CABG	1388	91/2	Leape et al. 1993
1993	PTCA	1306	58/4	Hilborne et al. 1993
1993	Coronary angiography	1355	76/4	Bernstein et al. 1993a
1993	Hysterectomy	642	59/16	Bernstein et al. 1993b

Source: Modified from Leape (1992).

Abbreviations: CABG, coronary artery bypass graft; PTCA, percutaneous transluminal coronary angioplasty.

ness rather than the inappropriateness of care. One very significant challenge lies in the choice of methodology (and, therefore, of evidence) that is used to assess appropriateness and to ground clinical recommendations. The RAND research uses a modified Delphi form of consensus methodology: a procedure is deemed appropriate (its anticipated benefits for a particular patient scenario exceed its anticipated harms) if it was judged by a nine-member panel of physician experts to be within a specified median range with no more than two physicians rating the procedure as inappropriate. Although panelists' evaluations were based in part on a synthesis of the relevant literature, only 10 percent of the available studies on coronary angiography, gastrointestinal endoscopy, carotid endarterectomy, and CABG were based on randomized controlled clinical trials (RCTs). Moreover, according to the RAND researchers, they made "no attempt . . . to score the quality of journals, articles, or [article] contents" that formed the empirical basis for their recommendations (Fink et al. 1987). A weakness of the RAND method, therefore, is that it is not based on an explicit link between recommendations and the quality of supporting evidence (Woolf 1992). Despite its more formal approach to consensus development, this method is also perceived to be limited by its fundamental reliance on opinion. On this point, advocates of the RCT as the proper method for determinations of appropriateness suggest that expert opinion may often be inaccurate and thus compromise initiation of needed controlled clinical trials. Richard Peto, a critic of the RAND methodology, has suggested that RAND consensus recommendations against the efficacy of carotid endarterectomy "had a major chilling effect" on efforts to determine the procedure's efficacy through an RCT (cited in Cotton 1993).

The RCT also has its shortcomings, however, as a method for determining appropriateness. RCTs do not provide evidence for a procedure's effectiveness (the level of benefit achievable under ordinary clinical conditions), but, rather, for its possible efficacy (the level of benefit achievable under ideal clinical conditions). RCTs are often extremely expensive, and their results may not be available for years. Moreover, many RCTs use a narrow range of outcomes measures and ignore quality-of-life and other health status variables that are relevant in health care decision-making (Guadagnoli and McNeil 1994). In addition, there are ethical challenges to randomization itself.

The RAND-modified Delphi approach and randomized trials represent only two in a growing number of sophisticated methods proposed to

assess clinical appropriateness. The Agency for Health Care Policy and Research (AHCPR) PORT studies, for example, provide a hybrid, evidence-based approach that also accounts for expert assessments. Meta-analysis is a method for analyzing a range of evidence from across multiple studies. The types and quality of evidence required to establish appropriateness are the subjects of ongoing debate. As Eddy (1993) has pointed out, the resolution of these important methodological questions will establish a standard for determining what works in medical practice, and this, in turn, will have an enormous influence not only on the reimbursement decisions of third-party payers, but also on the approval process for research protocols that seek a determination of efficacy through an RCT. In the next section we look at a number of proposals that have been offered on the evaluation and ranking of evidence.

A second challenge that accompanies the stress on appropriateness of care is the medical profession's historical tendency to view appropriateness in strictly clinical terms. This perspective is illustrated and advocated in a letter to the editor of the *New England Journal of Medicine*. "The patient's attitude," the author says, "should have no influence on what the physician advises as appropriate therapy for the patient's illness. The . . . risks to benefit ratio, as stated by the physician, should be the only consideration when making a therapeutic decision" (Wortzman 1979). Incongruities between patient and physician perceptions of appropriateness, however, have been demonstrated in a number of studies. In a comparison of risk preferences regarding prostatectomy, for example, Barry and colleagues (1988) have shown that patients presented with an analysis of risks and benefits tend to decide in favor of the procedure less often than their physicians. Similar differences have been observed in physician and patient perceptions of the appropriateness of laryngectomy as a treatment for throat cancer (McNeil, Weichselbaum, and Pauker 1980).

These findings point to the fact that the concept of appropriateness is fundamentally evaluative. It is variously informed by the norms and values of science, of medicine, of individuals, and of society. Determinations of the appropriateness of a procedure therefore should not be regarded simply, or even primarily, as an evidentiary problem (since evidentiary issues are themselves value laden), but rather as a problem of values assessment. Appropriateness research has become increasingly attentive to this fact and has expanded its decision models to include not only clinical values (such as changes in physiological function) but pa-

tient values as well. In what follows, we will specify the concept of appropriateness from the clinical point of view, the perspective of the individual patient, and the perspective of society.

Clinical Assessment: Correlating Evidence and Efficacy

In the context of patient care, the recommendation of a procedure is based on at least four interrelated clinical factors: (1) the patient's clinical profile; (2) the physician's (and team's) skill; (3) the quality of the evidence supporting a procedure; and (4) the procedure's clinical benefit/harm ratio, understood in terms of the empirical evidence that is available on the intervention's clinical efficacy and effectiveness. We use the term "benefit" advisedly here, cognizant of the fact that ultimately it is the patient who determines whether a specific clinical effect is, in fact, beneficial—a subject we take up in the section on "desirability" below. From the physician's point of view, however, the *clinical* benefit/harm ratio represents one way of assessing the potential effects of a procedure in terms of such empirical and population-based variables as morbidity, mortality, and quality-of-life indicators. If an intervention is inconsistent with the patient's clinical presentation or if the physician or team is insufficiently skilled to carry out the intervention (e.g., has a high complication rate), then the intervention is *de facto* inappropriate and should not be recommended.

Evidentiary judgments are understood in terms of degree and thus can be roughly classified as a basis for clinical recommendations. Proposed rules of evidence for such classifications have recently come from a number of quarters. As reported by Sackett (1989), participants in a conference on the use of antithrombotic therapy have distinguished five levels of evidence:

- I. large RCTs with clear-cut results (and low risk of error) [low false positive errors and high power]
- II. small RCTs with uncertain results (and moderate to high risk of error) [high false positive errors and/or low power]
- III. nonrandomized, concurrent cohort comparisons
- IV. nonrandomized, historical cohort comparisons
- V. case series without controls

Participants subsequently classified clinical recommendation into three grades, depending on their supporting levels of evidence. Grade A recommendations are those supported by level I evidence. Those in grade B are supported by evidence level II. Grade C recommendations are based on evidence from levels III, IV, or V.

This classificatory framework parallels the one put forward by the U.S. Preventive Services Task Force in its evaluation of the effectiveness of clinical preventive services (U.S. Preventive Services Task Force 1989). The task force proposed a hierarchical ranking of evidentiary quality that gave greater weight to study designs whose methodology made them less subject to bias and inferential error. In descending order, they are well-designed RCTs followed by nonrandomized controlled trials, cohort studies, case control studies, comparisons between time and place, uncontrolled experiments, and, finally, descriptive studies and expert opinion. Using this hierarchy, recommendations were graded on a five-point scale.

In what follows we provide a framework for correlating evidence and efficacy as a basis for clinical recommendations and informed consent. Our hierarchical ranking of evidence is similar to the rankings discussed above. In addition, we use this framework to suggest terminological distinctions that can be used in place of the ambiguous terms "appropriate" and "necessary," which, for the purposes of this article, we regard as synonymous (Kahan et al. 1994).

The *possibly beneficial* intervention is one thought to be preponderantly beneficial, but only on the basis either of evidence provided largely by case reports or of uncontrolled clinical impressions or uncontrolled studies. In other words, such inferences about the benefit of an intervention are rationally plausible but have only been subjected to the weakest empirical scrutiny. An example would be radical prostatectomy for well-differentiated, localized prostate cancer in patients less than 75 years old (Fleming et al. 1993). Cases where uncontrolled studies provide the evidence for an intervention's efficacy include photorefractive keratectomy as a surgical treatment for myopia (Gartry, Kerr, and Marshall 1992; Salz et al. 1993) and CABG in asymptomatic patients or patients with ischemia or mild angina/ischemia with three-vessel disease, no severe left anterior descending (LAD) stenosis, and normal left ventricular (LV) function (American College of Cardiology/American Heart Association 1991). It should be made clear to the patient that procedures in this category are based on the weakest empirical support.

Indicated and *highly indicated* interventions are supported by proportionally superior empirical evidence as to their benefit. The evidence supporting an indicated intervention should be derived from at least one randomized, controlled clinical trial, from a broad meta-analysis of sound studies (L'Abbé, Detsky, and O'Rourke 1987), or from a systematic consensus methodology or outcomes review like the PORT projects described above (Clinton 1991). An example would be conventional lumbar laminectomy/discectomy for patients with uncomplicated herniated discs (DATTA Report 1990, 1991; Weber 1983).

The *highly indicated* procedure is one whose benefit has been unequivocally established through either definitive or replicated RCTs. Additionally, the highly indicated intervention is one whose expected benefits have been clearly shown to exceed the anticipated benefits of alternative therapies. Examples include carotid endarterectomy for patients with transient ischemic attacks (TIAs) and greater than 70 percent stenosis (North American Symptomatic Carotid Endarterectomy Trial Collaborators [NASCET] 1991; Mayberg et al. 1991) or CABG for patients with left main artery disease and chronic, stable class III angina (American College of Cardiology/AHA . . . 1991).

Just as the evidence regarding efficacy and effectiveness reveals the relative benefits of an intervention, so too does it reveal the relative harms. Accordingly, we recommend replacing the overly broad terms *inappropriate* and *unnecessary* with the more specific designations described below. Like the previous distinctions regarding benefit, these terms have as their primary reference point the degree of empirical evidence substantiating a procedure's benefit/harm ratio. Again, these distinctions become relevant only when a procedure is consistent with a patient's clinical presentation and when the physician (and team) are adequately skilled. If *these* conditions are not met, then a procedure is contraindicated regardless of the strength of evidence for its general efficacy. An important exception is the performance of procedures by physicians in training. Society allows for supervised intervention by medical students and residents and accepts the attendant risks because the future availability of medical care is viewed as a significant social good. Whether or not these risks are equitably distributed remains an important question for public policy.

When a procedure is thought to be preponderantly harmful based only on anecdotal evidence, case reports, or uncontrolled studies, the procedure should be termed *possibly harmful*. Examples would include lumbar fusion for back pain (Turner et al. 1992; Franklin et al. 1994) and decalcification for degenerative aortic stenosis (Freeman et al. 1990).

These lower levels of evidence also support the *equivocal* intervention: one whose ratio of benefit to harm is roughly equal. An equivocal designation indicates, above all, the need for more study on the particular intervention and its clinical uses. Examples include carotid endarterectomy for TIAs and 30 to 70 percent stenosis (NASCET 1991; European Carotid Surgery Trialists' Collaborative Group [ECST] 1991) or for asymptomatic bruit (Barnett and Haines 1993; Bornstein and Norris 1993).

Important ethical limits on human subjects research prevent the performance of rigorous trials to prove that a drug or procedure's burdens are indeed disproportionate to its benefits. Thus the evidentiary requirements to establish a procedure as *contraindicated* will not be as demanding as those placed on determinations of care that is indicated or highly indicated. Accordingly, when good evidence from controlled trials, from a broad meta-analysis of studies, or from a systematic consensus methodology or outcomes review reveals that the harms outweigh the benefits of a procedure, that procedure should be termed "contraindicated." Examples of contraindicated procedures are extracranial-intracranial bypass for stroke prevention (Barnett et al. 1985), CABG in patients with mild stable angina and single-vessel disease (American College of Cardiology/AHA 1991), radical prostatectomy for well-differentiated localized cancer in patients older than 75 years (Lu-Yao et al. 1993; Wennberg et al. 1988), and carotid endarterectomy in patients with TIAs and less than 30 percent stenosis (ECST 1991). When superior evidence has established that the possibility of benefit and harm from an intervention is roughly equivalent—that is, when there is clearly no favorable benefit/harm ratio—that intervention should be classed as *nonindicated*.

Figure 1 represents the distinctions just described as they appear along the spectrum of evidence and the spectrum of benefit and harm that can be associated with a procedure. As represented by the vertical axis, evidence may be weak or strong depending on study methodology and scope. As represented by the horizontal axis, the efficacy (and/or effectiveness) of an intervention may extend from clearly preponderant benefit to clearly preponderant harm. Although this axis represents only magnitudes of benefit and harm, probability assessments are also relevant to the judgment process. The probabilities of benefit and harm are data that should be elicited from the available evidence and transmitted to the patient in the informed consent process (McNeil et al. 1982; Eddy 1990b).

The shaded area in the figure encompasses interventions that, because of insufficient evidence, should be identified as unproven. When any such intervention is contemplated, the fact that it is unproven should be

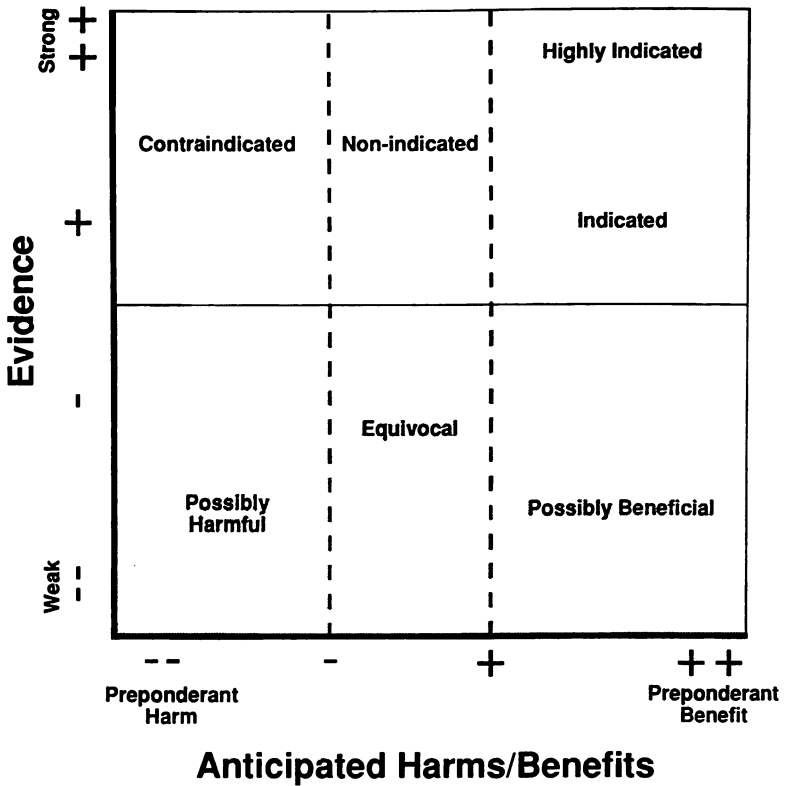


FIG. 1. Clinical assessment: correlating evidence and efficacy. Shaded area denotes classes of interventions whose efficacy and effectiveness are as yet unproven.

made clear to the patient in the informed consent process. Ideally, interventions of unproved benefit should be undertaken, when ethically permissible, only within the context of a research protocol. Figure 2 is a sample plotting of the interventions described above.

The classifications represented in Figure 1 provide a degree of specificity not captured by the broad terms "medically necessary" or "clinically appropriate." As such, they can provide a coherent and meaningful vocabulary for medical practitioners and patients as they deliberate on the right course of treatment. These classifications can also serve as a useful analytic framework for practice guidelines. Specialty societies might, for example, use this figure as a model to identify, in an ongoing way, the correlation between evidence and the benefit/harm ratio for the inter-

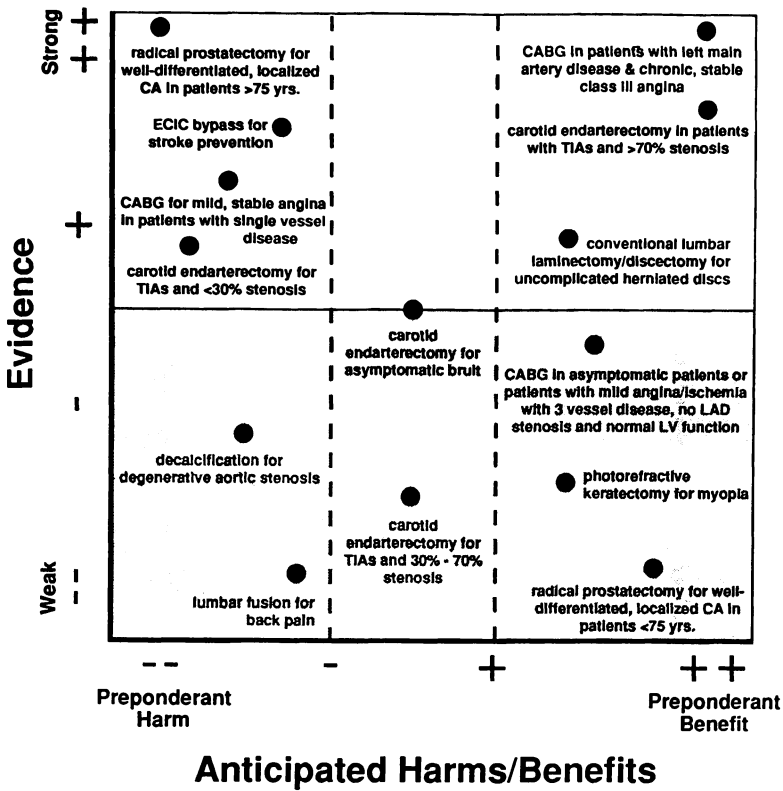


FIG. 2. A sample plotting of interventions. Shaded area denotes classes of interventions whose efficacy and effectiveness are as yet unproven.

ventions particular to their specialties. This information would ideally be supplemented by broader outcomes data on the comparative efficacy and effectiveness of alternative treatments. Further, as evolving informed consent law recognizes their availability, these data will also increasingly be regarded as material information that *must* be disclosed to patients (Hatlie 1993).

In the following sections we discuss two different ways in which patient considerations should inform determinations about the legitimacy of an intervention. In the context of outcomes research, data about the *average* patient should be incorporated into patient-centered outcomes measures. In the therapeutic context, considerations of the *individual* patient should determine the *desirability* of an intervention.

Efficacy, Effectiveness, and Patient-Centered Outcomes Measures

Because clinical trials are considered the optimal source of information for medical decision-making, the measures used within these trials to determine a procedure's efficacy and effectiveness must be as broad as possible and should include well-standardized, patient-centered quality-of-life and health status measures like anxiety, impact on role and social function, recuperation time, and days lost from work (Brook 1989; Tarlov et al. 1989; Tarlov 1992). To date, as Geigle, Brook, and others point out, the narrow focus on mortality and morbidity or change in a physiological variable, with its attendant omission of quality-of-life measures, has left the patient largely on the periphery of most outcomes assessment and outcome-oriented quality measurement (Geigle and Jones 1990; Brook and Kamberg 1987; Lehr and Strosberg 1991). By incorporating patient-centered health status measures in their overall assessment of efficacy and effectiveness, recent studies have been able to provide a more comprehensive range of relevant information as the basis for informed health care decision-making (McNeil, Weichselbaum, and Pauker 1981; Ware et al. 1981; Lohr 1988; Hollenberg et al. 1991).

Assessments of the quality of evidence supporting a procedure will be based on the degree to which they have included a broad array of clinical and patient-centered outcomes measures such as survival rate, states of physical, emotional, and social health, quality of well-being, and patient satisfaction. In the best-case scenario, all of this information would be available to individual patients and physicians in the clinical setting and would provide a comprehensive basis for informed consent or refusal. In other words, the more that technology and therapeutic assessment includes patient-centered considerations, the more relevant the data on efficacy and effectiveness will be to actual patients. Eddy has suggested that a distinction be made between such "intermediate outcomes" as test results or biological or physiological indicators and the ultimate "health outcomes" of a procedure. He argues that health outcomes—effects that patients experience and care about like pain, anxiety, death, disfigurement, and disability—should be the primary focus of informed medical decision-making (Eddy 1990a). When interventions have as one of their main purposes the improvement of quality of life, the "appropriateness" or "necessity" of a treatment cannot be rigidly defined according to non-patient-centered criteria (Barry et al. 1988).

Empirical information is only one component in the determination of an intervention's legitimacy. A further element is the intervention's *desirability* as judged by the informed patient.

The Patient's Perspective and the Desirability of Care

It is generally accepted that the appropriateness or necessity of an intervention is to be understood in terms of potential benefits and harms that it offers to the patient. Given this, the notion of appropriateness or necessity must take into account not only the *clinical* benefits and harms—understood both narrowly in terms of morbidity and mortality and more broadly in terms of overall quality of life for the “average” patient—but also the *relevance* of clinical and “nonclinical” benefits and harms for the *individual patient* in the context of medical decision-making. This includes individual quality-of-life decisions and individual assessments of acceptable risk and cost. Of course, as outcomes measures become more patient centered, the notion of a “clinical” benefit will be correspondingly enlarged. Nevertheless, in the context of care, an individual patient's values, risk preferences, and financial and social circumstances introduce considerations that are essential to the decision-making process and that cannot be captured by aggregate data about health outcomes. In short, it is the *individual patient* who determines the ultimate desirability of an intervention for him- or herself. For instance, although a surgery might be regarded as highly indicated from the point of view of relative medical risks and benefits, it might also be regarded as undesirable by a patient who either is risk averse or rejects anticipated medical benefit in favor of other, more highly valued goals (Pauly 1979; McNeil, Weichselbaum, and Pauker 1978; Danis et al. 1988). Albert Einstein, for example, refused surgery for an aortic aneurysm (which ultimately killed him) because he was committed to a life based on simplicity. He judged the recommended surgery as undesirable because increased longevity was, in his view, outweighed by the inconvenience of the intervention (Gary A. Chase: personal communication, December 12, 1994). Another example of an intervention that could be deemed indicated yet undesirable is a laryngectomy for a throat cancer patient who is willing to opt for quality of life over survival because she values the continued

use of her voice over longevity without normal speech (McNeil, Weichselbaum, and Pauker 1981).

In order to honor the values that the *individual patient* brings to the decision-making process, we recommend that an intervention be termed *desirable* if it is freely accepted by the informed patient or valid surrogate.

Interpreting Appropriateness and Futility in the Clinical Context: Some Ethical Considerations

Avoiding Conflict through the Process of Informed Consent/Refusal

The distinctions we have made between a procedure's clinical merit and its desirability to the patient are an attempt to make explicit the fact that the broad terms "necessary" and "appropriate" by definition imply the endorsement of some goal. The same is true for the concept of futility, a notion that has gained wide currency in recent medical literature (Pellegrino 1993). "Necessary for what and whom?"; "appropriate to what and whose end?"; "futile in the achievement of what and whose therapeutic aim?" are questions that must be answered if the terms are to be made meaningful (Truog, Brett, and Frader 1992; Youngner 1990). The informed consent process should involve, therefore, a broad discussion of the therapeutic goals that may differently motivate patients and physicians and lead them to diverse interpretations of the "necessity," "appropriateness," and "futility" of different interventions. The explicit and ongoing identification of the goals of therapy not only facilitates understanding and collaboration in the patient-physician relationship but also lessens the likelihood of conflicting assumptions about care. As explicated in a Yale-New Haven Hospital policy on do-not-resuscitate decisions (Committee on Policy . . . 1983), patient care may have as its goal (1) the achievement of cure or remission; (2) the maintenance of biological function; or (3) the maximization of comfort. Typically, in the routine practice of medicine, it is tacitly understood that the objectives of cure or remission are primary. Even here, however, explicit discussion is necessary to clarify the risks that the patient is willing to accept in the achievement of possible benefit. When, because of advanced illness, the goal of cure is unattainable, the "appropriateness" and desirability of various in-

terventions will need to be reexamined in light of the newly understood objectives of care (American College of Obstetricians and Gynecologists 1995).

A key ingredient of the process of informed consent is a discussion between physician and patient of both the relative benefits and harms of different treatment options and the quality of evidence that supports these judgments. Insofar as it allows the rough location of a procedure within the related spectra of evidence and benefits and harms, figure 1 can serve as a useful tool in the decision-making process. Recent technological innovations may further enhance the information provided to patients. For example, on the basis of evidence that surgeons and patients often interpret the need for surgery differently, one program has supplemented physician-patient discourse with an interactive videodisc (Wennberg et al. 1988; Randall 1993).

This process of informed consent is the means by which a recommended intervention becomes meaningful to the patient as "desirable" or "undesirable." The concept of futility, like the concept of appropriateness, can be variously interpreted by the clinician, the patient, or by cost-containment experts. As a result, comparable clarity is required when making use of this term in the context of health care decision-making.

Ethical Guides for Conflict Resolution

Cases will at times arise when the informed consent process results in conflict regarding the goals of therapy. In the recent *Wanglie* case,¹ for example, a patient's family insisted on the desirability of care that the providers believed to be futile (viz. without clinical benefit). More commonly, in cases concerning the withdrawal and withholding of treatment, a patient may refuse undesired care that physicians or hospitals deem necessary. Some ethical guidelines are useful when reflecting on conflicts of this sort.

First, based on the fundamental liberty of patients as persons, health care providers have no unilateral right to implement a decision about what they consider necessary or unnecessary in the context of individual patient care. Rather, their legitimate domain is the assessment of clinical

¹*In re Helga Wanglie*, 4th J.D. (Distr. Ct., Probate Ct. Div.) PX-91-283, Minn., Hennepin Co.

efficacy, and the probable effects of alternative modes of therapy versus no therapy, and the discussion of this information and their recommendations with the patient or a valid surrogate. In the context of nonemergency care, the physician's authority to perform an intervention derives from the patient's (or surrogate's) consent to a procedure that he or she deems desirable.

Second, the profession of medicine is guided by its moral commitment to avoid harm to patients and to aid them in maintaining or improving their health. On this basis, clinicians have a presumptive obligation not to provide treatments that are unproven or contraindicated. In addition, each physician is a moral agent who cannot be compelled to violate his or her personal moral convictions. For these reasons, a patient's demand for unproven or contraindicated care that he or she deems desirable is not sufficient to impose upon providers an obligation to provide that care.

The Societal Dimension of Appropriateness and Futility

As growing concern over both access to health care and escalating health care costs has made clear, individual and professional aims are not pursued within a vacuum. They are pursued within the context of health care institutions—hospitals, HMOs, third party insurance—and within the larger societal context where the diverse aims of health care, education, defense, and environmental protection, among other sectors, must compete for limited social resources. In the institutional context, the issues of appropriateness and futility are embedded within the practices of gatekeeping and utilization review, and they have found their way into new hospital policies on futility (Meyer 1993). In the larger context of health policy, the societal meanings of appropriateness and futility are manifested in debates about “global budgets,” a “decent” minimum level of health care (President's Commission . . . 1992), “essential” and “nonessential” services (Eddy 1991), and health care rationing (U.S. Congress 1991; Strosberg et al. 1992). For better or worse, health care decisions are tied to the marketplace and the wider context of social policy.

It is our task as citizens to determine what goals are and are not worth pursuing given the forces of human need and market economics (Pellegrino 1986). This determination is fundamentally based on the

values we hold rather than on any "facts" that can be supplied to us by scientific investigation. As Truog and colleagues have pointed out, "The decision that certain goals are not worth pursuing in the context of health care is best seen as involving a conflict of values rather than a question of futility" (Truog, Brett, and Frader 1992). To underscore that point, we would recommend, with Tomlinson and Brody (1990), that when treatment decisions are based on considerations of social distribution, rationing, or cost containment, the appropriately specific term would be the *cost worthiness* of an outcome, an intervention, or a likelihood of success, rather than its necessity or futility. This terminology makes explicit the economic values upon which a decision rests. By contrast, the term "futile" should be reserved for those procedures that, from the clinical point of view, offer no possibility of controlling or reversing the course of illness or improving a patient's desired quality of life. The term "futile," in other words, should be used to refer to interventions that are deemed clinically nonbeneficial, independent of financial considerations. As Tomlinson and Brody (1990) observe, in this era of cost containment, providers must maintain a clear distinction between the clinically nonbeneficial operation and one that is simply judged not cost worthy. Candor in this regard, at the public and individual level, will strengthen public trust in physicians and provider institutions and will help us to understand the human and economic consequences of our choices and our institutional structures.

As Daniel Callahan (1991) has observed, "Life was easier when we thought medical 'necessity' and 'futility' were scientifically discoverable." Knowing now that necessity and futility are not so much discovered as *decided upon* on the basis of certain value commitments (of physicians, patients, and citizens), we must turn our attention to the difficult task of establishing a public standard that determines the boundaries of what society may offer in the arena of health care. Callahan suggests that such a standard should incorporate at least six elements:

1. medical need defined in some general way
2. the efficacy of available treatments in meeting that need
3. the comparative costs and benefits of those treatments
4. the necessity of setting health care priorities
5. a political process capable of making the combined medical and moral judgments that will unavoidably be encountered along the way

6. the stimulation of public and professional debate on the substantive content of the moral judgments."

Until we have begun to gain some clarity on these issues, we would be wise to refrain from the uncritical use of the ambiguous and often misleading terms "necessary," "unnecessary," "appropriate," "inappropriate," and "futile." The use of more specific terminology by providers, patients, and policy makers may prevent mistaken assumptions, enhance the informed consent process, and advance the public discourse.

Unnecessary Surgery—Reprise

The shift in focus from "unnecessary surgery" to "appropriateness research" is evidence of a new era in the evaluation of health services, one that has brought increased scrutiny of the effectiveness and efficacy of medical and surgical practices. Arnold Relman (1988) has dubbed this the era of "assessment and accountability." As concerns about cost containment provide incentives to cut back on expensive services, it has become increasingly clear that the problem of *underuse* may surpass overuse or unnecessary treatment as a serious risk to patients. Unlike research on unnecessary surgery, the focus on appropriateness has the advantage of being able to address the quality of patient care in terms of both the *under*-utilization and the *over*-utilization of services. Insofar as "appropriateness research" lays open the issue of unmet health care needs, it may also provide some insight into the goal of universal access to care.

References

- American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures. 1991. Guidelines and Indications for Coronary Artery Bypass Graft Surgery by the Subcommittee on Coronary Artery Bypass Graft Surgery. *Journal of the American College of Cardiology* 17:543-89.
- American College of Obstetricians and Gynecologists. 1995. End-of-Life Decision Making: Understanding the Goals of Care (ACOG committee opinion 156). Washington, D.C.
- Barnett, H.J.M., and S.J. Haines. 1993. Carotid Endarterectomy for

- Asymptomatic Carotid Stenosis. *New England Journal of Medicine* 328:276-79.
- Barnett, H.J.M., S.J. Peerless, et al. 1985. Failure of Extracranial-Intracranial Arterial Bypass to Reduce the Risk of Ischemic Stroke. *New England Journal of Medicine* 313:1191-200.
- Barry, M.J., A.G. Mulley, F.J. Fowler, and J.E. Wennberg. 1988. Watchful Waiting vs. Immediate Transurethral Resection for Symptomatic Prostatism: The Importance of Patients' Preferences. *Journal of the American Medical Association* 259:3010-17.
- Bernstein, S.J., L.H. Hilborne, L.L. Leape, M.E. Fiske, R.E. Park, C.J. Kamberg, et al. 1993a. The Appropriateness of the Use of Coronary Angiography in New York State. *Journal of the American Medical Association* 269:766-9.
- Bernstein, S.J., E.A. McGlynn, L.L. Siu, C.P. Roth, M.J. Sherwood, J.W. Keeseey, et al. 1993b. The Appropriateness of Hysterectomy. *Journal of the American Medical Association* 269:2398-402.
- Berwick, D.M. 1989. Health Services Research and Quality of Care. *Medical Care* 27:763-71.
- Bornstein, N.M., and J.W. Norris. 1993. Evolution and Management of Asymptomatic Carotid Stenosis. *Cerebrovascular and Brain Metabolism Review* 5:301-13.
- Brook, R.H. 1989. Health Services Research: Is It Good for You and Me? *Academic Medicine* 64:124-30.
- 1991. Quality of Care: Do We Care? *Annals of Internal Medicine* 115:486-90.
- Brook, R.H., and C.J. Kamberg. 1987. General Health Status Outcome Measurement: A Commentary on Measuring Functional Status. *Journal of Chronic Disease* 40 (suppl. 1):131S-136S.
- Callahan, D. 1991. Medical Futility, Medical Necessity: The Problem Without a Name. *Hastings Center Report* 21:30-35.
- Chassin, M.R., J. Kosecoff, R.E. Park, C.M. Winslow, K.L. Kahn, N.J. Merrick, et al. 1987a. Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? *Journal of the American Medical Association* 258:2533-7.
- Chassin, M.R., J. Kosecoff, D.H. Solomon, and R.H. Brook. 1987b. How Coronary Angiography Is Used. Clinical Determinants of Appropriateness. *Journal of the American Medical Association* 258:2543-7.
- Clinton, J.J. 1991. From the Agency for Health Care Policy and Research. *Journal of the American Medical Association* 266:2057.
- Committee on Policy for DNR Decisions, Yale-New Haven Hospital. 1983. Report on Do Not Resuscitate Orders. *Connecticut Medicine* 7:477-83.

- Cotton, P. 1993. Determining More Good Than Harm Is Not Easy. *Journal of the American Medical Association* 270:153.
- Danis, M., M.S. Gerrity, et al. 1988. A Comparison of Patient, Family and Physician Assessments of the Value of Medical Intensive Care. *Critical Care Medicine* 16:594-600.
- DATTA Report. 1990. Laminectomy and Microlaminectomy for Treatment of Lumbar Disc Herniation. *Journal of the American Medical Association* 264:1469-72.
- 1991. Reassessment of Automated Percutaneous Lumbar Discectomy for Herniated Discs. *Journal of the American Medical Association* 265:2122-3,2125.
- Eddy, D.M. 1984. Variations in Physician Practice: The Role of Uncertainty. *Health Affairs* 3:74-81.
- 1990a. Anatomy of a Decision. *Journal of the American Medical Association* 263:441-3.
- 1990b. The Challenge. *Journal of the American Medical Association* 263:287.
- 1991. What Care Is "Essential"? What Services Are "Basic"? *Journal of the American Medical Association* 265:782-8.
- 1993. Three Battles to Watch in the 1990s. *Journal of the American Medical Association* 270:520-6.
- European Carotid Surgery Trialists' Collaborative Group. 1991. MRC European Carotid Surgery Trial: Interim Results for Symptomatic Patients with Severe (70-99%) or with Mild (0-29%) Carotid Stenosis. *Lancet* 337:1235-43.
- Fink, A., R.H. Brook, J. Kosecoff, et al. 1987. The Sufficiency of the Clinical Literature for Learning about the Appropriate Use of Six Medical and Surgical Procedures. *Western Journal of Medicine* 147:609-15.
- Fleming, C., J.H. Wasson, P.C. Albertsen, M.J. Barry, and J.E. Wennberg. 1993. A Decision Analysis of Alternative Treatment Strategies for Clinically Localized Prostate Cancer. *Journal of the American Medical Association* 269:2650-58.
- Franklin, G.M., J. Haug, N. Heyer, S.P. McKeefrey, and J. Picciano. 1994. Outcome of Lumbar Fusion in Washington State Worker's Compensation. *Spine* 19:1897-903.
- Freeman, W.K., H.V. Schaff, T.A. Orszulak, and A.J. Tajik. 1990. Ultrasonic Aortic Valve Decalcification: Serial Doppler Echocardiographic Follow-Up. *Journal of the American College of Cardiology* 16:623-30.
- Gartry, D.S., M.G. Kerr, and J. Marshall. 1992. Excimer Laser Photorefractive Keratectomy. *Ophthalmology* 99:1209-19.

- Geigle, R., and S.B. Jones. 1990. Outcomes Measurement: A Report from the Front. *Inquiry* 27:7-13.
- Gray, D., J.R. Hampton, S. Bernstein, J. Kosecoff, and R.H. Brook. 1990. Audit of Coronary Angiography and Bypass Surgery. *Lancet* 335:1317-20.
- Guadagnoli, E., and B.J. McNeil. 1994. Outcome Research: Hope for the Future or the Latest Rage? *Inquiry* 31:14-24.
- Hatlie, M.J. 1993. Climbing the Learning Curve: New Technologies, Emerging Obligations. *Journal of the American Medical Association* 270:1364-5.
- Hilborne, L.H., L.L. Leape, S.J. Bernstein, R.E. Park, M.E. Fiske, C.J. Kamberg, et al. 1993. The Appropriateness of the Use of Percutaneous Transluminal Coronary Angioplasty in New York State. *Journal of the American Medical Association* 269:761-5.
- Hollenberg, N.K., M. Testa, et al. 1991. Quality of Life as a Therapeutic Endpoint: An Analysis of Therapeutic Trials in Hypertension. *Drug Safety* 6:83-9.
- Kahan, J.P., S.J. Bernstein, L.L. Leape, L.H. Hilborne, R.E. Park, L. Parker, et al. 1994. Measuring the Necessity of Medical Procedures. *Medical Care* 32:357-65.
- L'Abbé, K.A., A.S. Detsky, and K. O'Rourke. 1987. Meta-Analysis in Clinical Research. *Annals of Internal Medicine* 107:224-33.
- Leape, L.L. 1992. Unnecessary Surgery. *Annual Review of Public Health* 13:374.
- Leape, L.L., L.H. Hilborne, R.H. Park S.J. Bernstein, C.J. Kamberg, M. Sherwood, et al. 1993. The Appropriateness of Use of Coronary Artery Bypass Graft Surgery in New York State. *Journal of the American Medical Association* 269:753-60.
- Lehr, H. and M. Strosberg. 1991. Quality Improvement in Health Care: Is the Patient Still Left Out? *Quarterly Review of Bioethics* 17:326-9.
- Lohr, K.N. 1988. Outcome Measurement: Concepts and Questions. *Inquiry* 25:37-50.
- Lu-Yao, G.L., D. McLerran, J. Wasson, and J.E. Wennberg. 1993. An Assessment of Radical Prostatectomy: Time Trends, Geographic Variation, and Outcomes. *Journal of the American Medical Association* 269:2633-6.
- Marwick, C. 1993. Federal Agency Focuses on Outcomes Research. *Journal of the American Medical Association* 270:164-5.
- Mayberg, M.R., S.E. Wilson, F. Yatsu, et al. 1991. Carotid Endarterectomy and Prevention of Cerebral Ischemia in Symptomatic Carotid Stenosis. *Journal of the American Medical Association* 266:3289-94.
- McNeil, B.J., R. Weichselbaum, and S.G. Pauker. 1978. Fallacy of the

- Five-Year Survival in Lung Cancer. *New England Journal of Medicine* 299:1397-1401.
- . 1981. Speech and Survival: Trade-Offs Between Quality and Quantity of Life in Laryngeal Cancer. *New England Journal of Medicine* 305:982-7.
- McNeil, B.J., S. G. Pauker, et al. 1982. On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine* 306: 1259-62.
- Meyer, H. 1993. Cost-Conscious Hospitals Set Futile Care Rules. *American Medical News* 3(June 28):20.
- North American Symptomatic Carotid Endarterectomy Trial Collaborators. 1991. Beneficial Effect of Carotid Endarterectomy in Symptomatic Patients with High-Grade Carotid Stenosis. *New England Journal of Medicine* 325:445-53.
- Office of Technology Assessment. 1983. *The Impact of Randomized Clinical Trials on Health Policy and Medical Practice*. Washington, D.C.: U.S. Government Printing Office.
- Park, R.E., A. Fink, R.H. Brook, M.R. Chassin, et al. 1986. Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures. *American Journal of Public Health* 76:766-72.
- Pauly, M.V. 1979. What Is Unnecessary Surgery? *Milbank Quarterly* 57: 95-117.
- Pellegrino, E.D. 1986. Rationing Health Care: The Ethics of Medical Gatekeeping. *Journal of Contemporary Health Law and Policy* 2(Spring):23.
- . 1993. Ethics. *Journal of the American Medical Association* 270: 202-3.
- President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. 1983. *Securing Access to Health Care*. Vol. 1: *Report*. Washington, D.C.: Government Printing Office.
- Randall, T. 1993. Producers of Videodisc Programs Strive to Expand Patient Role in Medical Decision Making Process. *Journal of the American Medical Association* 270:160-2.
- Relman, A. 1988. Assessment and Accountability: The Third Revolution in Medical Care. *New England Journal of Medicine* 319:1220-2.
- Sackett, D.L. 1989. Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest* 95 (suppl. 2):2S-4S.
- Salz, J.J., E. Maguen, A.B. Nesburn, C. Warren, J.I. Macy, J.D. Hofbauer, et al. 1993. A Two-Year Experience with Excimer Laser Photorefractive Keratectomy for Myopia. *Ophthalmology* 100:873-82.
- Stroman, D.F. 1979. *The Quick Knife: Unnecessary Surgery*. New York: Kennikat.

- Strosberg, M.A., J.M. Weiner, R. Baker, and I.A. Fein (Eds.). 1992. *Rationing America's Medical Care: The Oregon Plan and Beyond*. Washington, D.C.: Brookings Institution.
- Tarlov, A.R. 1992. Outcomes Assessment and Quality of Life in Patients with Immunodeficiency Virus Infection. *Annals of Internal Medicine* 116:166-7.
- Tarlov, A.R., J.E. Ware, S. Greenfield, E.C. Nelson, E. Perrin, and M. Zubkoff. 1989. The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care. *Journal of the American Medical Association* 262:925-30.
- Tomlinson, T., and H. Brody. 1990. Futility and Ethics of Resuscitation. *Journal of the American Medical Association* 264:1276-80.
- Truog, R.D., A.S. Brett, and J. Frader J. 1992. The Problem with Medical Futility. *New England Journal of Medicine* 326:1560-4.
- Turner, J.A., M. Ersek, L. Herron, J. Haselkorn J., D. Kent, et al. 1992. Patient Outcomes after Lumbar Spinal Infusion. *Journal of the American Medical Association* 268:907-11.
- U.S. Congress. 1991. *Who Lives, Who Dies, Who Decides: The Ethics of Health Care Rationing* (Report of the Senate Special Committee on Aging). Washington, D.C.: U.S. Government Printing Office.
- U.S. Preventive Services Task Force. 1989. *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions*. Baltimore: Williams & Wilkins,
- Ware, J.E. Jr., R.H. Brook, A.R. Davies, and K.N. Lohr. 1981. Choosing Measures of Health Status for Individuals in General Populations. *American Journal of Public Health* 71:620-5.
- Weber, H. 1983. Lumbar Disc Herniation: A Controlled Prospective Study with Ten Years of Observation. *Spine* 8:131-40.
- Wennberg, J.E., and A. Gittelsohn. 1973. Small Area Variations in Health Care Delivery. *Science* 182:1102-8.
- Wennberg, J.E., B.A. Barnes, and M. Zubkoff. 1982. Professional Uncertainty and the Problem of Supplier-Induced Demand. *Social Science and Medicine* 16:811-24.
- Wennberg, J.E., A.G. Mulley, D. Hanley, et al. 1988. An Assessment of Prostatectomy for Benign Urinary Tract Obstruction: Geographic Variations and the Evaluation of Medical Care Outcomes. *Journal of the American Medical Association* 259:3027-30.
- Winslow, C.M., J.B. Kosecoff, M. Chassin, D.E. Kanouse, and R.H. Brook. 1988. The Appropriateness of Performing Coronary Artery Bypass Surgery. *Journal of the American Medical Association* 260: 505-9.
- Woolf S.H. 1992. Practice Guidelines, a New Reality in Medicine. *Archives of Internal Medicine* 152:946-52.

- Wortsman, J. 1979. Letter to the Editor. *New England Journal of Medicine* 300:928.
- Youngner, S.J. 1990. Futility in Context. *Journal of the American Medical Association* 264:1295-6.

Acknowledgments: We are indebted to Gary Chase, F. Daniel Davis, Ruth Faden, Bernard Gersh, John Griffith, Elaine Larson, Edmund Pellegrino, Susan Stocker, our colleagues at the Kennedy Institute of Ethics, and two anonymous reviewers for their insightful comments on earlier drafts of this article. We also thank Charles Buhs, Coleen MacNamara, and Elizabeth Wellner for their excellent technical support. This article was supported by a grant from the Charles E. Culpeper Foundation.

Address correspondence to: Virginia A. Sharpe, Ph.D., c/o Office of the Dean of Research and Graduate Education, Georgetown University Medical Center, 4000 Reservoir Road, N.W., Washington, D.C. 20007.