

What Is Unnecessary Surgery?

MARK V. PAULY

*Department of Economics, and Center for Health Services and Policy
Research, Northwestern University*

THERE APPEARS to be a growing trend toward public questioning of medical decisions. A leading element in this trend has been the recent discussion concerning unnecessary surgery in the United States. The prospect of greater government financing of medical care has accelerated this concern, which will undoubtedly extend to other medical therapies. While questioning is appropriate, the discussion so far has not really clarified the issues.

In this paper, I will argue that much of the confusion results from a failure of most medical or public health experts to state clearly what is meant by "unnecessary surgery." The problem is not, however, that the experts know what unnecessary surgery is, and have been unable or unwilling to communicate it, but, rather, that medicine as a discipline cannot generate either the conceptual apparatus or the complete information set needed to arrive at a general definition. I will further argue that some concepts that are familiar to economists, but not well known to most medical experts, will, *when combined with legitimate and important medical information*, permit a useful definition. Although there are some situations in which medical evidence alone may be sufficient to determine the undesirability of a procedure, the definition I will suggest is more appropriate because it is more general. It not only classifies such situations in an appropriate way, but also permits useful distinctions in cases in which a definition based only on medical information alone might be ambiguous or misleading. I will also speculate on the situations in which a medical definition alone is useful for purposes of determining policy, and suggest that they are relatively few compared to the situations in which my proposed definition is relevant.

The appropriateness of a concept of necessary care obviously depends on the use to which the concept is put. The concept of unnecessary surgery has been used primarily for public policy purposes, rather than for attempts to improve actual clinical performance. That is, conjectures about the extent of unnecessary surgery have been used as a rationale for suggested changes in physician reimbursement (away from fee-for-service), and in manpower policy (toward training fewer surgeons). My definition is intended to be useful, at least initially, in the public policy debate, although it may have relevance for practice as well. For this reason, I will discuss the applicability of this definition in the context of the congressional hearings on unnecessary surgery during the last few years. Finally, I will try to place the definition in the more general context of the relationship between private and social choice in medical care.

What Is Unnecessary?

Definition of Necessity

Presumably, the goal of medical care is to improve the well-being of individuals. If this premise is accepted, it follows that judgment of necessity must ultimately depend upon the effect of a procedure on patient well-being. Many medical activities provide some potential benefits to individuals, as well as some costs. Benefits may include improvement in mortality and morbidity, reduction in pain or discomfort, or greater peace of mind. Costs may represent the value of resources used, with value measured by the other goods the resources could have produced, or costs may involve possible complications or side effects that themselves reduce health.

To judge necessity, all of the costs or outcomes that matter to people must be weighted, or made comparable in some way. If a procedure costs \$X, yields some potential health benefits, has some positive probability of complications, and causes discomfort and worry for some period of time, then all of these outcomes need to have a value placed on them. Some of the outcomes will have positive values, and others negative ones. Because the outcomes usually will not be known beforehand with certainty, the definition should be in terms of the expected value of the benefits and costs. For instance, if an outcome would yield a benefit of x if it occurred,

but only has a 50–50 chance of occurring, its expected benefit would be 0.5x. If we call the outcomes with positive values “benefits,” and the outcomes with negative values “costs,” we can say that a procedure is defined as unnecessary if expected benefits fall short of expected costs. Put another way, a procedure is unnecessary if, on balance, it makes the individual worse off.¹ While such a definition may sound trivial, I would argue that it is not. In particular, it points our thinking toward the appropriate questions: What are the benefits? What are the costs? How should they be valued?

Evaluation of Benefits and Costs

Two things are required for evaluation of benefits and costs. First, it must be known, at least in a probabilistic sense, what the costs are, and the consequences of a set of activities. Second, the costs and consequences must be valued according to some common standard. The usual economic approach, which I shall adopt here, is to consider the individual consumer as the appropriate person to evaluate his or her well-being. But one should not think of an anxious, poorly informed patient as this relevant consumer. Instead, we wish to define necessity by referring to the choices of a potential patient *who has at least as much knowledge and understanding of possible costs and consequences as the physician*. This does not mean, it should be emphasized, that the patient is completely informed; the physician may not have complete information about all aspects of the patient’s condition. But it does mean that the patient has as much information on the procedure’s costs and consequences as is available, both beneficial and adverse. We call this individual the “fully informed consumer.” Then surgery, or any other procedure, would be judged to be necessary if the fully informed consumer would choose it, and unnecessary if the consumer would not choose it. In effect, this artificial consumer combines the physician’s information (or the hospital bookkeeper’s) about costs and consequences with information on personal preferences and income. Such a consumer chooses

¹Necessity is not, of course, the only aspect of a procedure that may be relevant. Using surgical rather than nonsurgical treatment for an illness distributes income toward surgeons, and the equity of this transfer may be questioned. However, if alternative therapy involves the use of other kinds of physicians, or imposes costs on consumers, it is unclear how the equity judgment should be made.

surgery if its benefits exceed its costs, when he is as cognizant of costs and benefits as is consistent with the present state of medical knowledge.

To the extent that the amount of information generated about a particular case can be varied, we simply use the same definition for the necessary amount of medical information. Given the physician's knowledge of the possible information to be generated by some test or laboratory procedure, and given the patient's values of the costs of tests and the benefits of outcomes, the appropriate level of medical information is thereby determined.²

The Preferences of the Informed User

The critical element in this definition is that the multitude of possible costs and consequences are all to be weighted, and the weights to be used should reflect the *preferences of the informed user*, not those of the attending physician, the academic medical expert, or the health planner or legislator.³ In practice, it may not be necessary for all actual consumers to be as well informed as this artificial consumer (though greater consumer information may in general be desirable). The kind of detailed cost-benefit calculus described here could probably not provide a practical method for day-to-day determination of the services needed. Rather, the choices of this ideal consumer serve as the method of defining what "necessity" means. The definition is also applicable whether or not choices are actually being made in a market setting in which there are positive money prices. I am therefore not suggesting that such a market must necessarily be used. Even with full-coverage national health insurance, the appropriate concept of necessity is the one given above; whatever

²An important assumption in this definition is that benefits from use accrue only to the direct user. This assumption may not always be legitimate; in some cases, there may be others who benefit. This is clearly so in the case of communicable disease, and, even for other conditions, others may be concerned about an individual's suffering or incapacity. We may deal with such concern by thinking of a subsidy representing the value these others place on the benefits that come from care. Then necessary surgery is what would be chosen by the fully informed individual when the cost to him is reduced by the amount of the subsidy.

³A similar point has recently been made by Richards (1978) except that his survey of patients (in which 91% of hysterectomy patients were pleased with the results) did not necessarily measure the preferences of *informed* consumers.

method is used to select the quantity of care to be provided, whether by alterations in reimbursement or in the availability of supply, the appropriate notion of necessity should still be based on selecting *as if* a fully informed consumer were choosing.

It should be emphasized that *both* clinical information and information about consumer preferences are needed to arrive at a satisfactory decision. The informed consumer's decision can be no better than the information that medicine has developed. Consequently, it is desirable to have better clinical information, if such information is not overly costly to obtain. However, that information alone is not generally sufficient to make a decision, and no expansion of the knowledge base will be sufficient unless it is supplemented by information on consumer preferences. At any point in time, medical knowledge is obviously given, and good decisions require this knowledge to be combined with information on patient preferences.

This definition and approach have the advantage of linking the technical problem of measurement to the persons and situations for which measurement is to be made. It emphasizes that "costs" and (especially) "benefits" are not abstractions reflecting planners' or experts' preferences and attitudes; they are not just debatable matters of opinion. Rather, the concepts represent a set of choices or preferred alternatives for consumers.

Defining and Measuring Benefits

The definition implies that, to determine whether a procedure is necessary or unnecessary, more information is required than just a list of possible medical outcomes and their probabilities. A set of weights for costs and consequences is also required. While physicians may have some information about these medical outcomes and probabilities, they are not formally instructed in ways of ascertaining the weights their patients attach to possible outcomes or to other goods that might have to be sacrificed to pay for the medical care. Practicing physicians may eventually develop a "feel" for these weights. (Indeed, the mark of successful clinicians may be their skill in drawing out these weights or preferences, as much as their knowledge of outcomes or their capability in predicting them.)⁴ Such

⁴Parker (1976) and Schwartz et al. (1973) have described applications of decision theory to medicine which attempt to implement what is described here.

skills are not a matter of particular professional expertise in a scientific or testable sense, nor are such skills likely to be especially characteristic of the academic physicians who are usually the "experts" in the debate. If the preferences of properly informed consumers could be ascertained, an additional exercise of merit might be to discover which kinds of physicians are best able to determine the weights for costs and consequences that their patients would actually want used.

Defining and Measuring Actual Costs

One final element in the definition that bears emphasis is that benefits are to be compared with *actual* costs. In the strictest sense, cost here should be defined as what economists call "opportunity cost"—the value of the inputs in their next best use, or the value of output foregone. What the patient typically pays for surgery might be expected to differ from this cost for two reasons.

First, insurance coverage, especially of in-hospital procedures, reduces what the patient pays well below the cost of the inputs, often to zero. But these costs must eventually be paid by someone, either taxpayer (in the case of Medicare and Medicaid) or insurance plan member (in the case of private health insurance.) Even with full information on the benefits of surgery, consumers who believe their costs to be what they pay at the point of use will tend to choose surgery that is worth less to them than its true costs, precisely because the cost users pay is ordinarily below the true opportunity cost. Individual consumers rationally ignore the small rise in premiums their use causes, but consumers collectively cannot do so; they must pay. So there is a theoretical presumption that some unnecessary surgery, in precisely the sense described above, may indeed occur. The critical point, however, is that at present the extent and distribution of this surgery are unknown, even to experts. The kind of information needed involves not only knowledge of the present situation, but also of what would happen if the user price were increased.

Some simple geometry may help to clarify the agreement. In Fig. 1, D_1 and D_2 represent two alternative possible demand or marginal benefit curves for fully informed consumers for some surgical procedure. If the price were set equal to marginal cost, the informed consumer would choose (on the average) Q_0 units. (It is

probably most useful to think of Q_0 as an average rate for a population, since surgery purchases are unusually infrequent.) If insurance reduces the price to zero, the fully informed consumer with demand D_1 will choose more surgery, or choose it more frequently. There will be some unnecessary surgery, surgery whose marginal benefit falls short of its marginal cost. In Fig. 1, this quantity would be Q_0Q_1 units, which could be relatively large. However, if the consumer's demand curve were as D_2 rather than D_1 , then the amount (and the importance) of unnecessary surgery would be relatively small, and might be disregarded.

A second reason for deviation between what patients pay and opportunity costs is that physician (and hospital) charges may not be related to opportunity costs. If, for example, surgeons would be will-

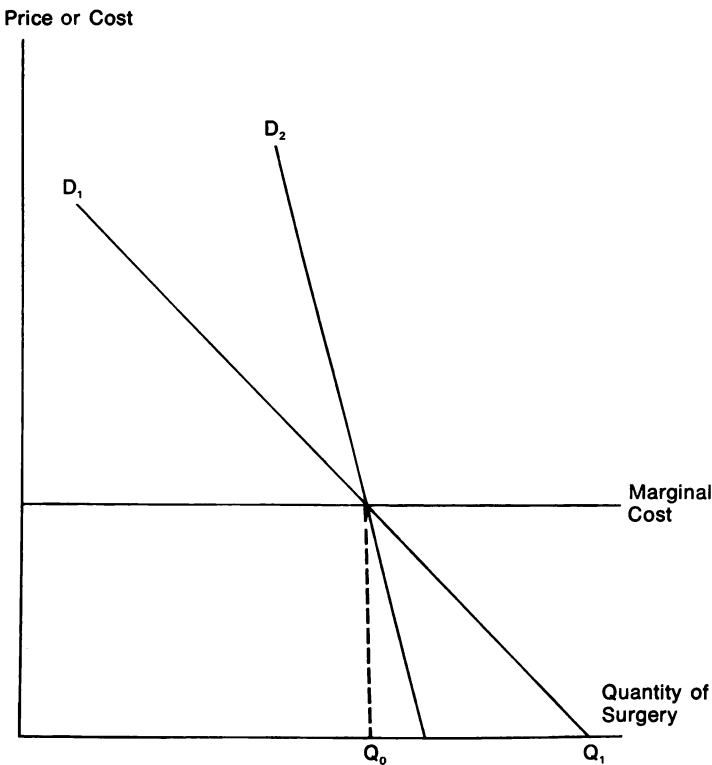


FIG. 1. Demand curves for surgery

ing to do more surgery at the current level of surgical fees than there is surgical work available, it can hardly be maintained that the surgeons' fee for performing surgery in a particular situation really represents the value of their time in other uses. The alternative uses of their time might be reading medical journals, providing primary care, or enjoying leisure, all of which would have less value than the surgical fee (either in terms of what consumers would pay or what surgeons would require). To the extent that surgical fees that are higher than opportunity costs discourage some otherwise well-informed people from seeking surgery, and induce them to use other forms of care or no care, there is too little surgery. Here again, exactly what the opportunity cost of surgical time is, how it is distributed over surgeons or surgical specialties, and how consumers might respond to fees that are closer to opportunity cost, is information that is unknown but required for proper empirical measurement of unnecessary surgery.

Using the Definition

There are two uses one can make of this definition—a positive one and a negative one. The positive use is to determine, in a better way than at present, what the consequences and weights are, or what the choice of the informed person would be, so as to arrive at a judgment about the necessity of procedures. I will discuss this in more detail at the end of the paper. But the more useful immediate application of this concept, and the one more relevant to the current debate over unnecessary surgery, is the negative one. This concept of necessity can be used to show that the definitions presently used to offer policy advice are seriously deficient, so much so that the advice is not legitimate. The point to be made is that, given our present state of knowledge, we cannot definitely (or even with a reasonable margin or error) say what is or is not necessary. Consequently, those who purport to do so are misleading their intended audiences.

Offering a brief for ignorance is not attractive, but it is important to recognize when knowledge is lacking if the best decisions are to be made. Greater damage can be done by decisions that are made on the assumption that knowledge or expertise is present than by decisions that are made with a frank acknowledgment of a present state of ignorance.

An Illustrative Example: Tonsillectomy and Adenoidectomy

Perhaps the best way to illustrate the usefulness of the definition I have proposed is to comment on a debate in the literature on one of the most commonly alleged forms of unnecessary surgery: tonsillectomy and adenoidectomy (T&A). My purpose here is not primarily to judge whether actual current rates of this surgery are desirable (although I will comment on that question), but rather to show how definitions can be used and misused.

Studies by McKee (1963a, 1963b) and Haggerty (1968) suggest the following consequences of T&A for the study population in Great Britain:

1. Reduction in the mean number of days of respiratory illness by 11 in the first year after T&A and 4 in the second year.
2. Reduction in the mean number of physician visits for respiratory illness by 1.5 in the first year and 0.7 in the second year.
3. Five days of convalescence associated with the procedure itself.
4. An increase in the risk of death of 1 to 3 per 10,000.

Other work (Roydhouse, 1969; Mawson, Adlington, and Evans, 1967) has produced results in the same range as these numbers, although there has been some more recent questioning of the benefit of the procedure (Roos, Roos, and Henteleff, 1977b). These findings suggest that there are some benefits from the procedure (whether directly or as a placebo), some potentially adverse health outcomes, and some explicit costs of the resources used up. If it were certain that there were no potentially positive benefits, but only possible adverse outcomes and/or costs of medical resources, then a judgment would be easy: T&A would obviously be “unnecessary.” But that does not appear to be the case here (nor is it generally the case); weights must be attached to each of the consequences of surgery, and not all weights will be negative.

A more subtle question is whether the benefits from T&A vary with some aspects of the patient’s history, such as the number of prior documented episodes of respiratory infection. This subject is currently under clinical investigation (Paradise et al., 1978), and some attempts have been made to provide answers from insurance records (Roos, Roos, and Henteleff, 1977b). The quantitative results are not definitive at present, but if it does become possible to observe

different expected benefit levels on the basis of prior history, then analysis of the necessity of a procedure will obviously be different for persons with different histories. In the absence of such definitive results, I will use a measure of average benefits in the following calculations.

The following calculations are intended to illustrate that it does matter which weights are used. For all of these consequences, it could seem reasonable to use as weights the values that parents place on costs and benefits. To convert them to a common measure, we will use the number of dollars (as a proxy for other goods) that parents would be willing to give up to obtain or avoid the particular outcome.

The McKee study (1963a) only followed the children for 2 years after tonsillectomy. Since respiratory illnesses become less frequent in later childhood in any case, it will be assumed that the benefits from the T&A continue at the second-year rate only for 3 more years (a total of 5 years post-tonsillectomy). The major benefits from the T&A, then, are: 1) a net reduction in days of illness of 22 over 5 years (27 days reduction less 5 days lost for the operation); and 2) a reduction in the number of doctor visits of 4.3 over 5 years.

Suppose 1975 data are used to estimate benefits and costs. If we value each day of illness at \$25 (what the parents might have to pay for a babysitter, perhaps, if both work, or what a middle income mother might pay to avoid caring for a sick child) and the doctor visits at \$30 (\$10 physician's fee and \$20 travel and waiting costs), the gross benefit is \$679.⁵

Against this must be set the costs: 1) doctor and hospital fees; and 2) an increase in the risk of death of 1 to 3 per 10,000 (probably an overstatement of the net risk, since a hospitalized or convalescent child may have a lower risk of death from other sources).

The mean stay for tonsillectomy patients under 15 years of age was about 1.8 days in Professional Activity Study (PAS) hospitals. If the cost of each hospital day is set at \$133, the 1975 average cost per adjusted patient day, hospital costs would amount to about \$240 per case.⁶ Physician bills would probably have amounted to about

⁵Strictly speaking, these future benefits should be discounted by some current rate of interest.

⁶For children treated on an outpatient basis, the cost would be lower.

\$150 (e.g., \$125 for the surgeon and an average of \$25 for the anesthesia, which would sometimes be included in the hospital bill when an anesthetist is used). Valuing life is always difficult, but a total value of perhaps \$200,000 is suggested by research on adults by Rosen and Thaler (1973). This yields a figure for increased risk of \$20 to \$60. Total costs, therefore, equal \$410 to \$450. The calculations ignore the psychological benefits to parents from reduced illness in their children and the psychological costs of surgery.

Necessity for T&A. I do not intend to argue that either these estimates of the effectiveness of the procedure or these measures of costs and benefits are perfectly accurate; indeed, my point is precisely that we do not at present have sufficient information to arrive at accurate numbers. Given these estimates, however, what can be said about the "necessity" of T&A? First, the part of the family's medical bills not covered by insurance would probably amount to an average of 10% of the total bills, or \$39. Adding on the estimates of the risk of death yields a cost to the family of \$59 to \$99, much less than the estimated benefits of \$679. From the viewpoint of the family, or the physician trying to do what is best for the family, the T&A seems clearly to be beneficial (in the cited studies). Even when the hospital and physician costs covered by insurance are added to get a measure of total social cost, however, T&A still yields benefits in excess of costs. The implication of the discussion is that, at present, no "expert" can reject the hypothesis that, on the average, T&A was beneficial in the cited studies to the extent of \$229 to \$269 per operation. Certainly, it is not possible to support North's (1968) contention (commenting on the Haggerty [1968] piece) that "the extent of total benefits seems trivial in relation to the costs." This would only be so if lower values are placed on sick days for children; whatever weights a pediatrician might assign, some or all parents may well have the kind of weights just mentioned.

It is more difficult to use these figures to determine whether unnecessary tonsillectomies are performed in the U.S. at the present time, because we have no comparable estimates of the average benefit for the U.S. population. (The British data are also not for a randomly selected sample, but since less than 10% of those referred for operation did not meet the selection criteria, one might suppose that the calculations provide a reasonable measure of average benefit for the United Kingdom.) Tonsillectomy rates in the U.K. are about

44% less than those in the U.S. (Bunker, 1970). It is unclear what effect these lower rates have on the measure of average benefit from the operation. But even if one makes the very strong assumption that the measure of average benefit should be reduced by 44%, the expected benefit from the procedure only falls to \$448, almost exactly equal to the estimated costs.⁷

Data from some sites in the U.S. are somewhat at odds with these figures, because they indicate that a large fraction of operations may not meet the selection criteria used in the British studies.⁸ Perhaps these differences are the result of alternative operational definitions of prior illness. In any case, these ambiguities in the data only serve to reinforce the conclusion that definitive evidence on the existence or magnitude of unnecessary surgery for tonsillitis is not available.

What Do We Really Know About Unnecessary Surgery?

With these observations as background, and using the conceptual definition developed above, let us look at some recent discussion of the evidence for unnecessary surgery. This evidence is of several kinds.

Variations in Surgery Rates and Unnecessary Surgery

There has been increasing documentation of a wide variation across geographic areas of the incidence of various elective surgical procedures. This is probably the "hardest" large-sample empirical evidence. In aggregate, these variations are often related to the presence of surgical specialists or hospital beds. This should not be surprising; other things being equal, and in the absence of excess capacity, more surgery will *necessarily* require more surgeons and more hospital beds. But the usual conjecture is that it is unlikely that

⁷Even though the "social benefit" from the procedure becomes less certain when this adjustment is made, there is still a substantial private benefit to the family with insurance.

⁸See, for example, the data from Seattle developed by Lo Gerfo and reported in Roos, Henteleff, and Roos (1977b).

the prevalence of the underlying conditions requiring surgery is subject to so great a variation (Roos, Roos, and Henteleff, 1977c). If surgery rates vary more widely than prevalence, then either there is unnecessary surgery, or some disease is not being appropriately treated.

The premise that the prevalence of conditions indicating elective surgery is not subject to such variation may well be correct. But the discussion above suggests that for many of these cases, given what we presently know, there is no basis for concluding that surgery is either superior or inferior to alternative forms of treatment. *We cannot at present, given the current state of knowledge, reject the hypothesis that surgery is approximately as good as nonsurgical forms of treatment for many of the symptoms that prompt such surgery.* Tonsillitis can be treated with antibiotics, menstrual bleeding with hormones or dilatation and curettage, and abdominal pain with observation. Although exact indifference is unlikely, we cannot prove that surgery is not a practical equivalent of these forms of treatment in managing the underlying condition.

Clearly, if surgical and nonsurgical methods of treatment are approximately equivalent, the "lack of necessity" of unnecessary surgery is likely to be small and the case for being concerned weak. Although hospitalization and surgery are expensive, the explicitness of these costs should not obscure the reality of other implicit costs of pain, uncertainty, discomfort, or inconvenience borne by patients who are not surgically treated. Of course, if surgery were known to be ineffective with regard to all of the characteristics that consumers value, then conclusions would be simple; but that kind of information is usually not present. Better clinical information would obviously *assist* in making better decisions, but such knowledge alone is not sufficient.

If the two methods of treatment are approximately equivalent, how can we explain the wide variations in their rate of use? One possible answer is that the method most likely to be used may depend upon the resources at hand. (How those resources become available is another, more complex question.) If an area is well-endowed with pediatricians, tonsillitis may well be treated with antibiotics and bed rest; but if there are few pediatricians and many surgeons (especially otolaryngologists), then we should not be surprised to find surgery chosen frequently. Indeed, even if actual physician bills and other costs are approximately the same between the

two methods, the opportunity cost of physician time—the value of a physician's time in the next best use—is not necessarily the same. Nor is the cost of this time necessarily measured by what a physician charges. If surgeons have a lot of spare time, on which they place little value, while pediatricians have no excess time, then society may lose less by having surgeons use their time to manage the condition surgically, as long as positive benefits are expected. This leaves open the questions of why bills do not reflect costs, and why consumers should enrich surgeons. But from the viewpoint of the best use of society's resources, it remains true that in areas where surgeons are more plentiful, other things being equal, there ought ideally to be more surgery.⁹

In summary, there may be significant variations in the (unknown) prevalence of conditions that prompt surgery, which can explain part of the wide variation in surgery rates. Even remaining variations need not imply widespread unnecessary surgery, if populations vary in the values they attach to the outcomes of surgery. Finally, even if the prevalence of conditions prompting surgery varies little, and even if patient "tastes" for surgery vary little, large differences in surgery rates still may be of little concern if the net benefits of surgical or nonsurgical forms of intervention are approximately equal. The actual number of procedures labeled "unnecessary" could be large, but the total amount of loss or waste could be slight. Because "necessity" is a concept that does not usually admit of degrees, it may not be most useful for judging the desirability of arrangements. What we really want to know is not how much unnecessary surgery there is, but rather how large a loss is imposed on the community. Widely differing rates of surgery, given our present knowledge, do not imply that there is widespread surgery that is really unnecessary. We cannot rule out the possibility on the basis of these figures, of course; all we can say is that the figures do not really help us very much.

⁹See Smallwood and Smith (1976) for a more extensive discussion of this point. Of course, at present there may be "too much" surgery almost everywhere, as judged by the definition of necessity. Even if surgeons are so abundant that their time has a zero opportunity cost, procedures with negative gross benefits, which harm the patient, obviously should not be done. Nevertheless, since in an ideal situation surgery should vary with surgeon availability, we cannot use an empirical finding of such a positive relationship to prove that there is unnecessary surgery being performed.

Surgical Rates in HMOs and Unnecessary Surgery

The second kind of evidence for unnecessary surgery is the difference in surgical rates for subscribers to prepaid groups (HMOs) compared to the general population. As distinct from the previous case, the problems here are mainly in the reliability of the numbers. If we could be sure that the populations were identical (or at least comparable) and that there was no measure of health or well-being by which HMO subscribers were worse than other persons in the population, then we would have to conclude that the differential did represent unnecessary surgery—costly surgery for which there was no benefit.

There are some adjustments that can be made to the data—adjustments for out-of-plan use and for observable differences in the population served—which in themselves can reduce the differential substantially. Klarman (1971) estimates that there is perhaps a 20% differential in total hospital admission rates after these adjustments are made, compared to a 50% differential before adjustment. To my knowledge, there is no documentation of the complete health experience of any HMO population, much less its overall well-being or level of satisfaction. Luft's recent survey (1978) of HMOs finds that their relative admission rates for elective surgery are not consistently lower than their rates for less discretionary admissions. Finally, because membership in an HMO is voluntary, we know that the populations are not identical—some people choose HMOs and others do not. Whether this difference matters for health care use is an unanswered question, as is whether these results will be repeated for all HMOs and all population groups.

Expert Opinion and Unnecessary Surgery

The third kind of evidence is based on what might be called "expert opinion." Expert opinion was the primary method used at the congressional hearings on unnecessary surgery in 1975 and 1977, and so a discussion of those hearings is appropriate here to indicate how definitions can be used and misused for policy purposes (U.S. House of Representatives, 1975, 1977b).

Surprisingly, none of the experts called before the House Subcommittee on Oversight and Investigations volunteered, or was asked, for a precise definition of "unnecessary surgery," other than

the tautological definition that "unnecessary" means "should not be done." While there was no discussion of definition, there was a discussion of definability during the first round of hearings. In perhaps the best statement, George Zuidema argued that the definition of "unnecessary" is very difficult, especially in those instances when experts differ. But by implication, and by his preference for the notion of "contraindicated" rather than "unnecessary," Zuidema seemed to favor a definition based on a consensus of physician experts. Obviously there are problems with such a definition, since for the kinds of surgery that vary widely in rates the experts do not agree on indications.

The "expert opinion" approach also appeared to be implicit in Sidney Wolfe's testimony. He began by talking about "unnecessary, criticizable, objectionable, should not be done, depending on what phrase you like," without recognizing that, given a sufficiently exhaustive search for an expert, there are many things that are criticizable or objectionable to someone. Wolfe, in fact, admitted that "unnecessary" is a word without distinct meaning: "I like to use the phrase 'unnecessary' because it is implied that this is being done in a circumstance when it should not be done."

But without saying how to tell when something should not be done, Wolfe went on to assert confidently that an observed increase in disc operations or hysterectomies must contain considerable amounts of unnecessary surgery. As an operational matter, he sometimes defined "unnecessary surgery" as the excess of the operation rate in fee-for-service medicine over that in prepaid group practice, a definition already discussed. But he also emphasized an expert opinion definition: the number of operations not recommended by a second consultant. Clearly this definition is very different from what was suggested above; it reflects in large part the differing judgments of two professionals, and that is all.

Second Opinion Programs and Unnecessary Surgery

The empirical evidence for the Subcommittee's subsequent conclusion in its initial report that there is a significant amount of unnecessary surgery was McCarthy and Widmer's study (1974) of a mandatory second-opinion program for a union health plan in New York City. These authors found that the second opinion was adverse to surgery for 16% to 18% of those patients referred for surgery.

These figures were used for nation-wide projections of the extent of unnecessary surgery (U.S. House of Representatives, 1977a).

There are a number of obvious technical objections to the conclusion that 18% of surgery is unnecessary. New York City has a much higher rate of surgical procedures than the country as a whole, so the projections were improper if the rate of unnecessary surgery is correlated with the overall surgery rate. It is also certainly possible that, had the consultants seen the entire set of patients, not just those referred for surgery, they might have *recommended* surgery for some of those not referred; how would such recommendations be classified?

However, the critical deficiency of using second opinion programs for estimating unnecessary surgery is that they are nothing more than opinion surveys. The results only tell us that, in the case of elective surgery, physicians have different opinions. What is striking about the results, viewed in this way, is that the extent of disagreement is so small; one might on the basis of other studies have expected more disagreement among physicians (Koran, 1975). Disagreement or not, the results are irrelevant, precisely because there is no presumption that either the referring physicians or the consultants attached the appropriate weights to all of the consequences and to the resource costs. That is, they did not necessarily make the same decision as would have been made by the fully informed consumer. They did not do this, because they could not do it; they would not in general have the necessary information on patient preferences or weights and resource costs.

What appears to have happened is that the pressure by politicians and others to render a judgment pushed experts beyond the boundaries of their expertise. Physicians do have specialized training in the treatment of disease, and they have some specialized knowledge on the relationship between alternative treatments and outcomes. This knowledge is necessary to judge the necessity or appropriateness of surgery, but it is not sufficient.¹⁰

¹⁰Even the evidence that might be gathered by medical audits or tissue committees, which determine whether or not disease or pathology was present, is not sufficient, because that information was not available at the time the decision was made to operate. An informed consumer may be willing to risk an operation prompted by symptoms that may be temporary if the consequence of postponement is some increased probability of a more serious adverse outcome.

What might have been meaningful would have been a comparison of actual surgery rates in two populations, one receiving information and opinions from a second consultant and the other not. The opinions should, of course, take the form of further information about the patient's condition and the consequences of surgery. Then one might argue that the experience of a better informed population, which more closely approximates the ideal customer, is being examined. The McCarthy and Widmer study did not examine such a comparison.

In a later study, however, McCarthy, Finkel, and Kamons (1978) did follow some of the patients who received a second opinion. Of patients not confirmed for surgery, almost one-third eventually had surgery anyway, so that only 11% of those referred for surgery actually did not have it. In later congressional testimony, McCarthy seemed to say that these 11% of referrals would have had "surplus surgery" in the absence of a second-opinion program (U.S. House of Representatives, 1977a). Even this statement is questionable, however, because without a control group one cannot know how many of the persons in the unconfirmed group would not have had surgery in the absence of a second opinion. McCarthy indicated that about 15% of the *confirmed* group—those who were twice advised to receive surgery—had not undergone the procedure on follow-up. It seems reasonable to suppose that a larger fraction of patients in the unconfirmed group might not have undergone the procedure. Without a control group, one cannot tell how many would not have had surgery, but these considerations suggest that, despite the apparent value of second opinion programs for the process of care, their value for judging the appropriateness of surgery is small.

Insurance and Unnecessary Surgery

A serious question is whether individual preferences, fully informed or not, should be taken into account in determining standards for surgery when third parties (government or private) pay the bill. Setting aside the kind of altruistic externality discussed above, do we not find that, when a collective pays for the care, society has a potentially different set of preferences that should be taken into account? A recent affirmative answer to this question was suggested by

Bunker, McPherson, and Henneman (1977): "If and when it is agreed to make 'necessary' medical care available to every citizen as a right, Society must decide what is necessary and unnecessary . . . At issue will be the allocation of public funds for a procedure when it appears to be more of a convenience or luxury than a necessity."

This position is not in accord with my definition. Bunker et al. appear to assume that "Society" is some distinct choice-making entity whose preferences ought to count. My definition views society as nothing more (or less) than the aggregate of its members, and bases measures of benefit on the value to those members. Changing the locus of choice may change the actual decision, but it does not change the ideal decision. The arbitrary division of procedures into luxuries and necessities only clouds the issue, since there is no reason why a set of persons who would individually choose a "luxury" should, when the decision is made collectively, suddenly decide only to purchase "necessities." The view that mortality reductions are necessities while life quality is a luxury is no more than an expression of personal preference.

What is true, of course, is that individuals presently choose care under insurance arrangements that conceal the true costs, whereas under National Health Insurance (NHI) the group of all consumer-taxpayers would have to pay the full cost. For this reason (and not because society has different preferences), the group might in effect choose to limit insurance reimbursement to procedures whose benefits equal or exceed their cost. But there is no reason to expect that marginal changes in mortality will somehow rank above relief of distress in this choice.

Moreover, the relevant criterion for choice is *still* the preferences of the informed consumer. The only effect of the move from individual to collective choice is an increase in the apparent cost to more closely approximate the true cost. This does, of course, raise an ambiguity in our evaluation of the current situation. If a consumer with full insurance obtains some positive (but small) benefit from some operation, it is rational to wish to receive the operation. Likewise, the physician should, if acting as the patient's true agent, recommend the operation, and, indeed, would not be acting in the patient's interest otherwise. Where the gross benefit is small (and this might be proxied by a difference of opinion among physicians in a second opinion program, as Enthoven (1978) has suggested), it is likely that benefit will fall short of costs. So,

paradoxically, the consumer may prefer a second opinion program that is linked to the payment of benefits not because it yields better information but because it may help to avoid operations that are of positive gross (but negative net) benefit.

Conclusions and Suggestions for Future Research

All of this negative counsel leads to a simple conclusion: We do not know, or even have a vague idea, whether unnecessary surgery, properly defined, is common or rare, nor can we provide any estimates that have any claims to validity. The definition that I have proposed is not at present fully operational. This does not mean that it is useless, however, or that it is not better than the alternative definitions implicit in the work I have discussed. While the tone must at present be negative, it need not be pessimistic.

Public Policy Implications

Knowing that we do not know means that we must select public policy based on the fact of ignorance. This conclusion suggests, for instance, that widespread alterations in manpower policy are unwarranted, as are attempts to encourage more surgery for cases of unmet needs. It suggests that individual physician and patient decisions might be improved—both by providing more information to patients on the usefulness of surgical treatment, and encouraging a greater flow of intelligence about surgery and surgeons in local communities. It also implies that financial incentives might be altered: insurance coverage on the patient side, the level of the fee received by the physician on the physician side. Because insurance typically covers all or most all of the surgical bill, even a patient fully informed about medical aspects would not be considering the true bill for the services to be received, much less the true cost. To the extent that fees for surgery exceed the value that surgeons place on other uses of their time, there is an incentive under fee-for-service to alter, by commission or omission, the information actually provided to patients. It is, however, a serious error to think that this is a defect of the fee-for-service system per se; rather the problem is that the fee has gotten too high. A reduction in surgical fees, coupled with an increase in reimbursement for surgical consultation, would surely lead to less surgery and more consultation.

Research Implications

It would appear that some progress could be made through research on measuring the preference weights, at least for average or representative kinds of consumers. It may be desirable, as well, to increase clinical information, since the consumer's decision (actual or ideal) can only be as good as the information on which it is based, but that information alone, however well developed, is not enough. Fully informed consumers could be sought out, although this might lead to the selection of persons unrepresentative of the population, such as Bunker and Brown's (1978) physicians' families. An alternative, which is more speculative but possibly more useful, would be to construct controlled experiments in which samples of consumers are given appropriate information and appropriate prices. Their choices could then be monitored, and the results of their choices (with suitable adjustments for individual error) used to guide collective choices and evaluations. There is sufficient precedent for such social experiments to suggest that they may eventually yield greater benefits than indiscriminate labeling of existing practice.

References

- Bunker, J. P. 1970. Surgical Manpower. *The New England Journal of Medicine* 282 (3): 135-144.
- , and Brown, B. E. 1974. The Physician-Patient as an Informed Consumer of Surgical Services. *The New England Journal of Medicine* 290 (19): 1051-1055.
- , McPherson, K., and Henneman, P. 1977. Elective Hysterectomy. In Bunker, J., Barnes, B., and Mosteller, F., eds., *Costs, Risks and Benefits of Surgery*. New York, Oxford University Press.
- Enthoven, A. C. 1978. Shattuck Lecture: Cutting Cost Without Cutting Quality of Care. *The New England Journal of Medicine* 298 (22): 1229-1238.
- Haggerty, R. 1968. Diagnosis and Treatment: Tonsils and Adenoids—A Problem Revisited. *Pediatrics* 41 (4): 815-817.
- Klarman, H. H. 1971. Analysis of the HMO Proposal. In *Health Maintenance Organizations: A Reconfiguration of the Health Services System*. Chicago, Ill.: Center for Health Administration Studies, University of Chicago.

- Koran, L. M. 1975. The Reliability of Clinical Methods, Data and Judgments. *The New England Journal of Medicine* 293 (13): 642-646.
- Luft, H. S. 1978. HMO's and Medical Costs: the Rhetoric and the Evidence. *The New England Journal of Medicine* 298 (24): 1336-1343.
- McCarthy, E. G., and Widmer, G. 1974. Effects of Screening by Consultants on Recommended Elective Surgical Procedures. *The New England Journal of Medicine* 291 (25): 1331-1335.
- , Finkel, M. L., and Kamons, A. S. 1978. Second Opinion Programs: A Vehicle for Cost Containment? In *Report of the National Commission on the Cost of Medical Care*, Vol. II. Chicago, Ill.: American Medical Association.
- McKee, W.J.E. 1963a. A Controlled Study of the Effects of Tonsillectomy and Adenoidectomy in Children. *British Journal of Preventive and Social Medicine* 17 (2): 49-69.
- , 1963b. The Part Played by Adenoidectomy in the Combined Operation of Tonsillectomy and Adenoidectomy. *British Journal of Preventive and Social Medicine* 17 (3): 133-140.
- Mawson, S. R., Adlington, P., and Evans, M. 1967. A Controlled Study of the Effects of Adenotonsillectomy in Children. *Journal of Laryngology* 81 (July): 777-789.
- North, A. F., Jr. 1968. An Epidemic Unchecked. *Pediatrics* 42 (4): 701-710.
- Paradise, J. L., Bluestone, C. D., Bachman, R. Z. et al. 1978. Limitation of Sore-Throat History as Indicator for Tonsillectomy. *The New England Journal of Medicine* 298 (8): 409-412.
- Parker, S. G. 1976. Coronary Artery Surgery: The Use of Decision Analysis. *Annals of Internal Medicine* 85 (July): 8-18.
- Richards, B. C. 1978. Hysterectomy: From Women to Women. *American Journal of Obstetrics and Gynecology* 131 (4): 446-449.
- Roos, L., Roos, N. and Henteleff, P. 1977a. Assessing the Impact of Tonsillectomies. Unpublished paper (June). Winnipeg, Manitoba: University of Manitoba.
- Roos, N., Henteleff, P., and Roos, L. 1977b. A New Audit Procedure Applied to an Old Question: Is the Frequency of T & A Justified? *Medical Care* 15 (1): 1-18.
- , Roos, L., and Henteleff, P. 1977c. Effective Surgical Rates—Do High Rates Mean Lower Surgical Standards? *The New England Journal of Medicine* 297 (7) 360-365.
- Rosen, S., and Thaler, R. 1973. The Value of Saving a Life: Evidence from the Labor Market. Working paper No. 7401 (December). Rochester, N.Y.: University of Rochester, Graduate School of Management.

- Roydhouse, N. 1969. A Controlled Study of Adenotonsillectomy. *Lancet* 7627 (November 1): 931-933.
- Schwartz, W. B., Gorry, G., Kassirar, J. et al. 1973. Decision Analysis and Clinical Judgment. *American Journal of Medicine* 55 (October): 459-584.
- Smallwood, D., and Smith, K. 1976. Optimal Treatment Decisions, Optimal Fee Schedules, and the Allocation of Medical Resources. Unpublished paper. Chicago, Ill.: Northwestern University.
- U. S. House of Representatives. 1975. *Getting Ready for National Health Insurance: Unnecessary Surgery. Hearings, July 15, 17, 18, and September 3, 1975.* Subcommittee on Oversight and Investigations, Committee on Interstate and Foreign Commerce. Washington, D.C.: U. S. Government Printing Office.
- . 1977a. *Cost and Quality of Health Care: Unnecessary Surgery.* Subcommittee on Oversight and Investigations. Washington, D.C.: U. S. Government Printing Office.
- . 1977b. *Quality of Surgical Care Hearings, April 25 and 29, May 2, and 9, 1977.* Subcommittee on Oversight and Investigations, Committee on Interstate and Foreign Commerce. Washington, D.C.: U. S. Government Printing Office.

Acknowledgment: The helpful comments of Edward F.X. Hughes, M.D., are acknowledged with gratitude.

Address correspondence to: Professor Mark V. Pauly, Department of Economics, College of Arts and Sciences, Northwestern University, Evanston, Illinois 60201.