

1-1-1992

The Public Policy Implications of Using Outcome Statistics

Mark V. Pauly

Follow this and additional works at: <https://brooklynworks.brooklaw.edu/blr>

Recommended Citation

Mark V. Pauly, *The Public Policy Implications of Using Outcome Statistics*, 58 Brook. L. Rev. 35 (1992).

Available at: <https://brooklynworks.brooklaw.edu/blr/vol58/iss1/3>

This Article is brought to you for free and open access by the Law Journals at BrooklynWorks. It has been accepted for inclusion in Brooklyn Law Review by an authorized editor of BrooklynWorks.

THE PUBLIC POLICY IMPLICATIONS OF USING OUTCOME STATISTICS

*Mark V. Pauly**

INTRODUCTION

The measurement of the outcomes of medical interventions is obviously a key ingredient in judging the effectiveness of those interventions. Surprisingly, however, a generally acceptable technology of measuring outcomes has never been developed. Instead, "quality" was traditionally quantified by measuring the performance of actions hypothesized to cause good outcomes, or by documenting the presence of structures hypothesized to cause the good actions that would cause good outcomes.

These halfway measures have increasingly come under criticism and a major effort has begun to measure outcomes as part of a larger attempt to determine which ways of rendering medical care really produce good outcomes. Exactly why the development of interest in outcomes has been so long in coming is not generally known. There doubtless are many causes, but surely one of them has been increasing concern with the cost of medical services. The cost of a medical service wonderfully focuses the discussion of the effectiveness of that service which is supposedly justified by the cost. It is probably no coincidence that the era of cost containment and the era of outcome measurement largely overlap.

When cost becomes an object of concern, agents of all types have a strong *desire* to measure the output of those services in an accurate and valid way. Here, as elsewhere in life, wishing does not necessarily make it so; the technology of output measurement is a long way from producing that unequivocal outcome measurement that payers, consumers, regulators, policy-

* Mark V. Pauly, Ph.D., is Chairman of the Health Care Systems Department and the Bendheim Professor of Insurance and Public Policy and Management at the Wharton School, University of Pennsylvania. He is also the Director of Research at the Leonard Davis Institute of Health Economics, University of Pennsylvania. Professor Pauly is nationally recognized as a leading health care economist for his writings on economics and health insurance.

makers and providers all crave so strongly.

The imperfection and dissatisfaction with outcome measurement technology are going to be a permanent state of affairs: there will *never* be perfect—or even near-perfect—measures of health outcomes. We may hope for the emergence of ways of measuring provider-specific outcomes with a reasonable degree of sensitivity and specificity,¹ but unless one uses an elastic definition of “reasonable,” I doubt that this will happen generally. The lack of a perfect measurement tool does not mean measures will not be developed for some diagnoses or procedures. Providers are going to be evaluated on the basis of outcome measures, but policymakers must cope with the fact that those measures will necessarily contain substantial imperfections. There will be a “signal” mixed in with the noise.² Spending more resources may raise the signal-to-noise ratio modestly, improving that ratio somewhat over time as new measurement technology becomes available. However, outcome measures will always need to be treated as aids to good guessing, as ways of improving the odds on a bet about which provider is better. The fundamental and irreducible uncertainty and imprecision in outcome measurement pose the biggest challenge to public policy.

In the face of this imperfect information, as luck would have it, some wrenching decisions will have to be made. These decisions involve cost containment to be sure, but they involve more than platitudes: they involve decisions about the rationing of care, the denial of benefits and substantial changes in the organization and structure of the health care system.

This Commentary critiques the myth of perfection in outcome data that has contributed both to the overselling of the concept of measuring provider outcomes in some settings and to a virulent reactive skepticism in others. The Commentary will also note that the public availability of outcome data is part of a challenge to some other dearly held myths about the health care system. Then this Commentary will turn to the question of appropriate public policy responses in the face of essentially imperfect data: what mistakes are to be avoided and what goals are

¹ Mark R. Chassin, M.D., M.P.P., M.P.H., *Quality of Care: Time To Act*, 266 JAMA 3472 (1991).

² That is, despite the imperfection, there will be valid information contained in the data.

to be (realistically) sought? Finally, it will examine the role of outcome data from a buyer perspective in a market-driven medical care system.

I. SOME FALLACIES

A. *Perfect Data*

Medical outcome data suffer from two kinds of imperfections. First, the outcomes measured are often incomplete or incorrectly measured. In-hospital mortality is presumably measured accurately, but avoidance of death in the hospital is surely not the primary objective most patients seek. Measurements of improvement in morbidity status, in functioning and in relief from pain or disability are all imperfect or incomplete. Measurements of degree of comfort and dignity associated with the treatment the patient receives barely exist. Second, initial or present severity is measured imperfectly. To determine whether care made a patient better, one needs to know both the final outcome and how sick the patient was before care was initiated. While there has been some development of measures of "initial severity," these measures never will capture all important aspects of the condition of every patient.³ Many of those aspects are initially unknown, even to the attending physician. Moreover, the physician will rarely record all the details of the patient's initial condition or other characteristics that may affect final outcomes.

These imperfections mean that a measured outcome for one hospital which is worse than some others can always be criticized: the provider can argue that "its patients are sicker" and that "it gets more subtle benefits than are reflected in raw mortality measures." Research suggests that there is some truth to these claims: some hospitals pinpointed as much worse than average in the Health Care Financing Administration (HCFA) mortality data fell back to just worse than average when detailed initial severity corrections were made.⁴

³ Robert W. Dubois, M.D., Ph.D., *Hospital Mortality as an Indicator of Quality*, in PROVIDING QUALITY CARE: THE CHALLENGE TO CLINICIANS 107 (Norbert Goldfield & David B. Nash eds., 1989).

⁴ S.D. Horn et al., *The Relationship Between Severity of Illness and Hospital Length of Stay and Mortality*, in MEDICAL CARE, April 1991 at 305-17; see also Dubois, *supra* note 3, at 122-24 (adjustments in variables from expected death rates when sever-

Even if reviewers of hospital outcomes could have access to the same information that the attending physicians had at the time of treatment, that is, to a complete and augmented medical record, the issue of "sicker patients" will still not be settled, at least as far as providers are concerned. An essential feature of public policy in this area is that outcome measures will be imperfect; they will sometimes indicate that a provider is worse than the provider really is, and sometimes suggest that the provider is better than the provider really is. Moreover, providers whose outcomes are worse than average will always disbelieve their bad records. In this sense, these measures are now and will remain unfair. Moreover, the search for a reasonably fair and reasonably accurate measure (depending on how one defines "reasonably") is bound to end in frustration.

B. *Feasible Uniform Quality*

If variation in hospital death rates can be linked, even probabilistically, to something other than variation in initial severity, some of those deaths must be preventable, as Robert Dubois has noted.⁵ That is, a given population of patients should expect to experience, on average, higher death rates in one hospital than in some others. Not all hospitals are equally good.

Not only are some hospitals and doctors more lethal than others, such differences may not be completely reducible. While changes in management practices including hiring policies, and in medical procedures and protocols, may reduce some of the outcome differences across providers, some differences will remain. Some physicians will always be more technically adept, more perceptive, or just plain smarter than others. Hospital managers also vary in skill. Not every hospital in town can have the best management.

Variation in skill or perceptiveness colors how providers interpret outcome data. The hospital or physician whose outcomes are worse than average will first review the process being used to provide care. Some deficiencies may be found that will be corrected. However, it is unlikely that quality will be fully brought up to the norm precisely because poor management often is

ity factored in).

⁵ Dubois, *supra* note 3, at 127-28.

caused by an inability to recognize bad process. That is, if the providers knew enough to understand (and accept) that their outcomes truly were inferior, they would have known enough to avoid those outcomes in the first place. This is why a "self-improvement" approach to raising quality will always fall short.

In most industries, variable quality—even when that variation intrinsically reflects the distribution of talents and motivations across individuals—is not a cause for concern. Within limits, lower quality care can be provided profitably in a market; it just sells for a lower price. The lower price in turn simply reflects the fact that the premium prices commanded by the best craftsperson or the best manager do not have to be paid to persons whose skills are average. In health care, however, the myth of "highest possible quality for all" is challenged by the common sense observation that it is simply impossible for all patients to have the highest possible quality. When outcome data did not exist, and when restrictions on advertising prevented invidious comparisons, it was possible to pay lip service to the myth that all doctors are equally good—even though health professionals knew that this myth was false. The challenge posed by the outcome movement raises the question of varying quality; some deaths or bad outcomes, literally speaking, are preventable, but cannot feasibly be prevented without mass expulsions of physicians and managers from the industry.

C. *No Tradeoffs*

Suppose, just for the moment, that we *could* measure outcomes perfectly. Such technology would be a big help in attempts to determine which care was effective and which was not: the technology could immediately help to eliminate types of care that do more harm than good, or that do no good at all, and it could be used to drive from the market providers unable or unwilling to provide care of positive net benefit. That is, all "dominated" actions would be eliminated, and only truly effective care would remain. However, there would still be a decision problem for buyers and policymakers: how effective must effective care be to justify its cost? When faced with the option of obtaining care of unequivocally positive, but small, benefit and high cost, should the buyer sacrifice the resources? Should public policy forbid the buyer from sacrificing resources, or require the buyer to do so? Better information will help make such an option

clearer, but it will not make the decision easier.

Let us return to the real world where information about the effectiveness of care provided by one provider (or one procedure) relative to another is imperfect. The raw numbers indicate that A is better than B, but one's confidence in this judgment is limited by statistical noise from sampling error and by knowledge of data imperfections. Of course, if the apparently better provider is cheaper than the apparently worse provider, there is less of an argument (although there is still some). Often, the most expensive hospital or the most expensive treatment is not the best. But we also know for a certainty that the cheapest hospital or the cheapest treatment in a town is not the best—but it is the cheapest. We also know that we will have to make tradeoffs under conditions of uncertainty. What must be traded off in this more realistic case is not only quality or good outcomes against cost, but also the degree of *certainty* about outcomes against cost. If the apparently better hospital is much farther away from home, or much more costly than some other hospital, but if I cannot tell for sure how much better it is, or even whether it is better, I will face a wrenching decision. As the odds that the hospital is better remain favorable, but decline, when do I decide to take a chance—to gamble that the cheaper or closer hospital is just as good? And, again, should public policy allow me to bet in any way I wish? This intrinsic tradeoff between uncertainty and "cost" (in monetary, convenience and amenity terms) is unavoidable, and is at the heart of the policy problem.

II. KEY POLICY QUESTIONS

How then should buyers use the emerging information, and what should public policy permit? I address five key questions:

- 1) Is picking the provider or the treatment with the best outcome statistics always the best bet?
- 2) What cautions, warnings, or prohibitions should be attached to the use of such outcome information?
- 3) Who should (or who should be permitted) to collect, analyze, publish, and disseminate outcome information?
- 4) How should employers and health care plans use such information?
- 5) How much effort and resources should be put into collecting accurate data?

A. *Picking Good Providers*

A buyer (who could be an individual patient, a benefits manager, or the medical director of a Health Maintenance Organization (HMO)) is trying to decide which hospital to use for a particular condition in a market area with many hospitals. Data have been collected on some dimensions of the outcomes of those hospitals. How, if at all, should those data be used?

There are two polar extremes. In Case A, all other information, about cost, quality, or uncertainty makes the choice a toss up; the hospitals are of about the same size, amenity, and convenience. All that is needed is a tie breaker. In Case B, one or the other provider appears superior on some other dimension—either cost (to the extent that is relevant), or some other measure of quality.

In Case A, the almost universal decision rule is obvious: when everything else is equal, pick the hospital with the best measured outcomes. The important point is that this rule should generally be followed regardless of the statistical significance attached to the comparison, and without regard to worries about the quality of the outcome or severity adjustment measures. Following such a rule means that the technique used both by HCFA and by some state data agencies in releasing their data—that of indicating as “significant” only those differences that meet a 0.05 criterion—will be misleading. Only a very conservative buyer would require that the chances that one hospital is better than another be greater than 19 in 20 before he or she would choose it; most buyers would be willing to go with the seller with the better record even without such strong reassurances.

However sensible this decision rule is for the buyer, the Case A decision presents problems for sellers: many hospitals (or physicians) will be shunned whose quality or outcomes, if the truth could be known, are as good or better than that of their apparently superior rivals. In this sense, using imperfect data is unfair to providers. There is some (but not much) consolation in noting that such imperfection affects firms in all other markets—the restaurant’s chef who is off on the day of the critic’s review, the teacher forced to cope with unruly students and the business’s delivery truck that breaks down through no fault of its own may also be scored as having lower quality than is the case. Markets do reward good quality, but only on average. The myth that “all hospitals and doctors are equally good,” pre-

served by inhibitions of various types on information to buyers, has protected providers for decades, but is now eroding. The consequence is that some good but unlucky providers will be grouped together with the less competent.

The policy question concerns the consequent reorganization of the industry in which some providers expand and others contract. The expanders, on average, will be of higher quality than the contractors, but there will be some good quality hospitals lumped in with the below average. There will be some wasted motion that will matter little over time, but in the short run will be of concern.

When would following the "pick the (measured) best rule in case of ties" be worse than the alternatives of picking the worst or picking at random? Following the rule *would* lead to a bad outcome if there was a *negative* correlation between measured outcomes and true quality, that is, if among two apparently similar hospitals, the one with the better statistics actually was, on average, worse than the one with the worse statistics. This possibility should not be ruled out, but is not very likely either.

Case B, in which one provider is apparently superior to another on other grounds, such as on the more traditional measures of structure or performance, or the basis of convenience or cost, will require more careful consideration by the buyer of the significance (statistical and practical) of measured outcome differences. Suppose the hospital most convenient to you, associated with your Preferred Provider Organization (PPO), and used by your favorite doctor, scores very poorly on outcome measures. The odds that the difference is due to chance alone are a little greater than one in 20, though no one can tell you the odds that the difference is due to mismeasurement of initial severity in the patient population generating the data. What should you do? The decision theory answer is simple but discomfoting: you will have to form some subjective estimate of the expected utility of the difference in outcomes *and* in the other dimensions, and choose the option with the highest expected utility. Less formally, you will have to decide on the tradeoff and whether the possible sacrifice in the chances of a good outcome are worth the savings in convenience cost, in money cost and in sticking with the doctor. What measures are needed to help you with this decision? Not the kinds of data that agencies typically publish, represented by tables with asterisks and statistical jargon. But it

is not obvious what represents a superior alternative.

Two policy strategies are worth considering. One represents an attempt (again) to explain tough choices under risk with imperfect data to a reluctant public. "Think of it like Atlantic City" is actually the best way to inform consumers (and the best time to explain Medicare mortality statistics might be on the New Jersey-bound chartered bus). But the political feasibility of an honest admission of the "good hunch" character of these choices remains troublesome.

While a frank admission of the risky nature of choice with outcome data is the best public policy in theory, policymakers in practice will doubtless find the development of such a policy challenging if not close to impossible. Government provided data that only changes the odds a little, but does not point the way to the unequivocally best (or target the unequivocally worst), will disappoint voters who would like the government to extract them from their painful dilemma. At the same time, the truly good quality provider indicted by a bad measure provided by the government—even the government agency that qualifies its numbers—will feel intensely wronged. There is no obvious solution to this problem. Of course, if outcome data were found to change buyer choices very little, then the information would not be worth the trouble. But if buyers do benefit from better knowledge, then policymakers should be advised to make the difficult choices.

The other strategy is to try to elicit, at some cost, the attitudes and preferences of a representative sample of buyers of various types of health care and to use those preferences to construct indicators that are intended to replicate their choices.⁶ Such a "ratings" approach will never be perfect, but it may be the best we can do.

B. *Cautions, Prohibitions, and Warnings*

I have already described the complex and dangerous nature of the choices that would need to be made as a result of outcome data. Providers live in mortal fear that these measures will be "misused." The next issue is what misuse means, how it can be

⁶ We are engaged in such an activity at the University of Pennsylvania for corporate benefits managers and medical directors, and are extending our efforts to union health plans.

prevented and whether it should be prevented.

The premise for this discussion has already been presented: using outcome data, either of the kind presently available or of the best kind we are very likely to get, will be difficult. Avoiding mistakes will be impossible and some understanding of how to interpret statistics will be necessary even to limit mistakes.

How should government (which itself is the source of many of the measures) behave? The safest course of action for the government is obviously to surround the data with a protective coat of qualifications and limitations. State data agencies have taken quite different approaches here: some go for a *caveat emptor* strategy, while others require, like a parent toward a child's potential escort, a much more explicit statement of intentions. A strategy of providing raw data with no help does not seem appropriate, but it is far from clear what sort of qualifications should be added. This is a new situation and we may well need to try some alternatives to see how buyers behave (and what they like best).

If such a situation were costless, there would obviously be no reason to oppose government provision of as much help as possible to interpret the data. In the real world, beyond requiring some basic cautions, there does not appear to be much government can or should do. The reason for this skepticism is that there will be a wide variety of potential uses for, and users of, the data. As already noted, the super-cautious approach to avoiding one type of error, that necessarily increases the likelihood of another type of error, is not what many buyers will want. If all buyers or users of the data had the same attitude toward risk, the same valuation of different dimensions of outcomes, and the same willingness to balance medical outcomes against cost or convenience, one might, after considerable effort, be able to develop a protocol or guide for use of the data. However, such uniformity does not generally exist and so the best action for government is to ensure that available data are as accurate as possible under the circumstances.

But is it not possible that consumers, patients or corporate buyers will make mistakes, where a mistake is defined as choosing a hospital that truly is lower quality than some other hospital? Would not a careful analysis of the use of outcome data compare the harms that arise from such mistakes with the potential gains in terms of better choices by others or greater effi-

ciency? The answer would appear to be affirmative, but there is an important paradox. There is not likely to be an external gold standard measure by which one hospital is definitely and definitively judged to be better or worse than another. If there were, the information used to generate that conclusion should itself be conveyed explicitly to consumers, patients and corporate buyers. Then, except for careless errors, the public should not make mistakes. The only mistake one can make using imperfect data is to rely on that data alone and ignore other ancillary information which, when combined with the data, will generate a reliable measure of quality. But then the solution is not to suppress the imperfect outcome data, but to combine them with the additional information that does lead to a correct conclusion.

We have the potential of observing a natural experiment in a new kind of market—the market for data interpretation. Both buyers and sellers of services have already put effort into assembling and trying to make sense of data and there is a potentially important service to be performed in collection, collation and interpretation of such information. So long as false and incorrect data are forbidden (as they should be), the further interpretation of the data will be a service for which a reasonably functioning market should emerge. At least, there seem to be no obvious impediments to such a market.

There is another possible consequence of imperfect provider specific data that might provoke a strategic buyer response. To the extent that providers have the power, they may try to avoid caring for patients who are suspected to be “sicker than average” in ways that are not taken into account by the outcome reporting system. Whether this behavior will be important is an open question. For instance, hospitals do not appear to have selected more profitable admissions to any great extent under Medicare. However, if some hospitals or doctors do channel sicker patients to other providers, one would expect sophisticated buyers to adjust for this behavior in their evaluations of outcome data. If private hospitals send sicker patients to municipal hospitals, this will make outcomes look better for the former and worse for the latter. Buyers should be able to detect this behavior and therefore will be less inclined to take some of the differences into account.

The main adverse consequence of imperfect outcome data in such a scenario will therefore be the channeling of patients to

providers who are not the most appropriate or the most convenient. Moreover, not all buyers will be sophisticated. Again, until we have more experience with outcome data, we do not know how severe a problem such sorting will pose.

A major purpose of providing outcome data is to allow buyers to make choices among sellers. The power to make choices is the opportunity to make mistakes and some errors are inevitable.

C. Who Should Collect, Publish and Analyze Data?

Information of all types is what economists call a "public good"—a good that a competitive market will supply in sufficiently small amounts because of the potential for free riding on others' efforts. Labeling a service a public good does *not* necessarily mean that a public entity needs to produce it, but it does mean that a legal requirement may be needed to provide data, and that some publicly financed subsidy for the collection and compilation of that data is justified.

In contrast to the need for an explicit public role in the collection of data, dissemination is a more complicated question. One issue is how the cost of disseminating information in some way varies with the number of users. On the one hand, if the cost is approximately proportionate to use, there is no need for public involvement. On the other hand, simple publication of the data might involve a single lump sum cost, with little additional cost for a user to be made aware of the results. In this case, some public role might be appropriate.

Analysis—the step between compilation and publication—poses more complex problems. One solution is a "two-track" strategy: publication of simple measures at public expense, and more complex analyses at private expense, especially for users who have special needs or desires. There is another loop in the process, however. How is a buyer of information to know which private seller is furnishing the best method of analysis? While there is in principle a need for "information about information" that might have some "public good" characteristics, the role appears to be minor. Seemingly a competitive process to supply information and analysis in various forms is both feasible and desirable. The likely structure of such an information industry is still a puzzle at the moment: entry appears easy, especially if the raw public data are made available at nominal

cost as they should be. Some firms might be able to develop special skills, but in general the market appears likely to remain both competitive and efficient.

D. *The Use of Information by Third Parties*

Use of information by the individual patient raises the issues of understanding and data availability. What about the use of information on outcomes by third party payers or by agents of those payers? Public insurers, private commercial insurers and the benefits departments of self-insured firms may all be interested in outcome measures for different providers. Such insurers may want this information either because they themselves intend to contract selectively with some providers, or because they want to be able to evaluate HMO and PPO arrangements that involve selective contracting. What motives would they be expected to bring to the use of data and should special limits or encouragements exist for them?

There is an important difference between such third parties and individual patients. Third parties are presumably interested in outcomes for *populations* of patients, while the individual patient cares only about the outcome for himself or herself. At a minimum, a third party may face a different kind of tradeoff, since a hospital that is best for one condition may not be best for all, but concentration of business may be necessary to get the most advantageous price. Moreover, the uncertainty about an outcome that might be large if applied to a single individual may average out in the experience of a reasonably large population of insureds.

The most serious puzzle about such third parties, however, is the question of their objectives. There is no doubt that some workers and unions fear the uses that their employers and insurers might make of outcome and cost data. On the other hand, employers might reasonably be expected to have as their objective the choice of arrangements that maximize benefits to employees, given the cost. After all, only Hawaiian employers are obligated to offer health benefits.⁷ Employers presumably pro-

⁷ Timothy Egan, *Hawaii Shows It Can Offer Health Insurance For All*, N.Y. TIMES, July 23, 1991, at A1. Hawaii is the only state to have near-universal health care. The program, initiated in 1974, requires employers to pay the cost of insurance premiums for any person working over 20 hours a week. The employee pays no more than 1.5% of

vide health benefits to attract and retain good workers at moderate wages.

At a minimum, employees would need to know what objectives are being pursued on their behalf. Theory suggests that health insurance benefits, which are offered to make working for a firm more attractive to the labor market, should be managed by firms to benefit employees; benefits managers ought to care about quality and outcomes as well as costs. Many do, but there is a potential danger for those employees that an employer is not trying to attract or retain. Some firms may not have a strong incentive to pay attention to outcomes, such as the company that is cutting back, the firm where benefits were extracted by unions and the corporation with employees in positions it wants to phase out.

E. *How Much Data?*

As in all other matters, spending more resources will, eventually, produce better results—in this case, better outcome measures. The key policy question is how good the outcome data need to be. The answer to that question depends, in turn, on the prior question of how the data are to be used.

For internal management of hospitals and other providers, the need for precision is likely to be different than for external valuation purposes. All judgments here are somewhat tentative, but I would speculate that internal management can proceed with data of poorer quality than can external valuation. Internal management, after all, collects much more data than simple outcome measures and obtains much more subjective information about performance. Outcome data are likely to be quite unsuited for internal management. One reason is that most of the outcomes measured in data sets that are comparable across hospitals are measures of serious adverse events which, thankfully, are rare, even for seriously ill patients: mortality, major morbidity, iatrogenic complications and readmissions for the same condition. The low frequency of these events means that the sample sizes at the level of the hospital floor, the physician or the short time period are generally too small to permit useful statistical

gross wages or $\frac{1}{2}$ the premium, whichever is less. *Id.* In 1990 Hawaii added the State Health Insurance Program that provides limited coverage to the 5% of the population not covered under the 1974 plan. *Id.*

comparisons. Since internal management also relies on analyzing variations within different parts of an organization, the parts need to be small enough to permit a sufficient number of comparisons—but then the sample sizes of patients are usually too small to permit the comparisons to have any strong statistical meaning. These observations suggest that outcome measures of the sort that public agencies collect may not be effectively used for internal management purposes. Instead, measures of process and the more frequently available measures of consumer satisfaction are likely to be used.

At the level of external purchasers, however, sample sizes may be large enough to generate sufficient statistical power for meaningful comparisons. The individual hospital cannot very effectively compare itself with other hospitals, but buyers can make that comparison. Buyers are also likely to lack the internal knowledge of process, of attitudes and of competence that management possesses. This means that valid external comparisons will be much more sensitive to imprecision of measurements, especially of initial severity. Moreover, most of the research activity in the next few years will be in the development of improved versions of severity measures. The key market question, however, is how much such refined measures are going to be worth to buyers.

For some buyers (with hope a small minority) there may be little interest in such information. Some benefits managers would find their lives a lot simpler if they could conclude that all hospitals in a town were equally good. Then they could just contract with the cheapest. While such an attitude is shortsighted, long-term vision in health care cost growth is a scarce commodity.

The most fundamental question is: what is it worth to a buyer to avoid a small mistake? In general, this value will be low: subtle unmeasured variations in initial severity mean that one might pick the second best hospital in town rather than the best, but to the buyer the difference is likely to be of little importance. There is a kind of Catch 22 here. One is only likely to suspect that more precise measurements of initial severity will make a major difference among hospitals where there are already obvious indicators of variations in unmeasured severity.

While subtle differences will not matter much to buyers, they may well matter to sellers. The best hospital that would

lose to the second best hospital which is almost as good has an incentive to make the truth known.

III. OUTCOME DATA AND THE MEDICAL ECONOMY

The key policy issue is not the quality or accuracy of the outcome measures to be developed. Some measures will eventually be developed, even though they will never be perfect. The key policy issue is *how* those measures will be used. Use, in turn, depends in large part on the rest of the delivery and financing system structure. I want to outline some thoughts on the likely use of outcome data in a competitive, market-based medical care delivery system and comment on the use of such data in a more centrally regulated system.

The ideal of competition involves much more than multiple producers of care trying to attract patients. Decades of evidence proves that competition alone, in the face of distorted incentives and imperfect information, will not produce an outcome that will make people happy. The ideal of a competitive market, as outlined by Enthoven, involves more than just many sellers.⁸ Such a market also involves *knowledgeable* buyers facing proper *financial* incentives. There is, it must be admitted, considerable debate, even among the proponents of competitive systems, about how they should be structured. There is general agreement, however, that buyers need to be conscious of both cost and quality as measured by outcomes.

Cost-quality consciousness is expected to produce three results, one much more emphasized and the other two less so. The most emphasized result is that competition, combined with cost-quality consciousness, should eliminate pure waste—where waste is defined as costly activities that do not yield any positive benefit. The reasoning is obvious: buyers always seek to avoid sellers who charge them more for the same outcomes as other sellers, or who give them inferior outcomes for the same price. With good outcome and cost data, and the option of choosing among a number of sellers, buyers will pick the seller with the least waste. Sellers who want to maximize profits, or who just want to survive financially, will find it in their interest to elimi-

⁸ ALAIN C. ENTHOVEN, HEALTH PLAN: THE ONLY PRACTICAL SOLUTION TO THE SOARING COST OF MEDICAL CARE 70-92 (1980).

nate waste. Overly costly ways of producing particular services and the supply of services that are costly but yield no benefit, will both be avoided.

A second result is more subtle: once the actions that are costly but yield no benefit are eliminated, buyers will need to compare the magnitude of services with positive benefits with their costs to judge whether the benefits are worth the cost. This choice will be difficult. It is not helpful to try to disguise the necessity of making this choice by trying to distinguish between services with "significant" or "more than marginal" benefits and other services. The key question is how large does the benefit have to be before it becomes significant. David C. Hadorn and Robert H. Brook try to capture this idea by distinguishing between effective care and "necessary" care, with necessary care (apparently) defined as care clearly demonstrated to provide "significant" net benefit.⁹ But other than tossing the question back to panels of supposed experts, these researchers provide no basis for distinguishing between "significant" positive net benefits and net benefits that are positive but "non-significant."¹⁰

The main point is that the emergence of accepted outcome measures will add a special urgency to the need to develop some way of making this tradeoff along the (almost) "flat of the curve." As long as the connection between care and outcomes is sufficiently vague, one can always pretend that some costly type of care is not "clearly" demonstrated to be effective and so might be ineffective. Better outcome measures and better linking of these measures to clinical actions will mean that the vagueness dodge will no longer work. But then we will be face-to-face with the policy question we all try hard to avoid: how do we say to someone, even to ourselves, that it is better to forego beneficial care because the care is not worth the cost in terms of non-medical goods and services that will have to be sacrificed?

The rising costs of medical care alone are pushing us willy-nilly toward having to confront the issue of rationing. When we do confront the rationing issue, outcome measures will be key. There is a real chance of backlash. The easiest way to avoid the difficult task of denying effective but costly care is to allege that

⁹ David C. Hadorn, M.D., M.A. & Robert H. Brook, M.D., Sc.D., *The Health Care Resource Allocation: Defining Our Terms*, 266 JAMA 3328, 3330 (Dec. 18, 1991).

¹⁰ *Id.* at 3329-30.

the data on effectiveness are no good. Most probably we will have to find a discourse that makes us comfortable with rationing—a discourse we have yet to find—before we can fully accept the use of outcome measures in that process.¹¹ We can see some of the alternatives in the discussion of Oregon Medicaid and the tendency of policymakers to fall back on opinions on what kinds of care are “good,” regardless of outcome evidence.¹²

The third difficult response to outcome data in a market setting will arise as markets do what they do best—cater to the *diverse* preferences of different buyers. It seems plausible that different consumers would want to make different choices in the cost-outcome tradeoff just described. To make those choices correctly—for example, to judge whether it is worth it to travel a longer distance, or pay more, to use a provider with better outcomes—the buyer will need to have information on outcomes. While I have in another setting advocated publicly financed assistance to enhance the ability of low income consumers to buy higher quality care, there will still be some residual variation, probably related to income, in the quality of care different consumers receive.¹³ Better outcome data will document this “inequality,” but will also facilitate such inequality. Here again, there is a real chance that the outcome measure will itself be blamed or criticized.

If different buyers do want different outcomes, the notion of a single set of standards or protocols for appropriate care will be challenged. Outcome measures may never become discriminating enough to define multiple guidelines, such as a high quality and

¹¹ *Id.* at 3328-29

¹² Oregon, in an attempt to extend Medicaid coverage to 120,000 poor citizens below the poverty level not currently eligible for coverage, planned to ration health services. State planners ranked 709 medical treatments in order of usefulness and cost and then estimated Medicaid could only provide coverage for the first 587. Michael Abramowitz, *Quashing the Oregon Plan*, WASH. POST, Aug. 11, 1992, at 27; Robert Pear, *Oregon's Experiment; Federal Rejection Blocks Health Care Rationing Plan*, N.Y. TIMES, Aug. 9, 1992, at A2. The conditions not covered include liver transplants for alcoholics and non-traditional measures to save low birthweight babies. Neal R. Peirce, *'AmBush' May Chill Health-Care Reform*, HOUSTON CHRON., Aug. 10, 1992, at A14. The plan was rejected by Secretary of Health and Human Services Louis W. Sullivan as violative of the Americans With Disabilities Act, as by definition, people with serious chronic illnesses are considered “disabled.” Michael Abramowitz, *Quashing the Oregon Plan*, WASH. POST, Aug. 11, 1992, at 27.

¹³ Mark V. Pauly et al., *A Plan for 'Responsible National Health Insurance,'* 10 Health Affairs 5 (Part 1, Spring 1991).

a not-so-high quality set of guidelines. But if such distinctions become possible, then the outcome measures that support them will come under more intense scrutiny.

Regulatory approaches to controlling costs and maintaining quality will almost surely not tolerate such diversity in either process or outcome. The difficult tradeoffs will still have to be made, however, and the regulator faced with accurate outcome data which show that some hospitals or doctors do not perform as well as others will be faced with a difficult choice: what does one say about the hospital that does not achieve as good an outcome as some other hospitals but is the only one convenient to a particular community? The outcome measures themselves will come under fire once they are taken seriously.

CONCLUSION

I have not dealt with the question of whether consumers, patients or corporate buyers can be trusted to use outcome measures, or whether they should be prevented from doing so. That is because, in my view, there is little chance of returning to the era in which "authorities" were trusted to assure medical services quality. In the role of consumer, or in the role of voter, citizens will be making choices based on outcomes.

