To Subscribe: https://fulfillment.healthaffairs.org

Health Affairs is published monthly by Project HOPE at 7500 Old Georgetown Road, Suite 600, Bethesda, MD 20814-6133. Copyright © by Project HOPE - The People-to-People Health Foundation. As provided by United States copyright law (Title 17, U.S. Code), no part of may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by information storage or retrieval systems, without prior written permission from the Publisher. All rights reserved.
Prologue: Knowing that wide variations in the use and cost of medical care exist in the delivery of health care is one thing, knowing what to do about them in a meaningful policy way is quite another. Bruce Vladeck, who has rapidly become a respected voice in health policy circles, addresses this question in this paper. Vladeck draws three conclusions, all of which reflect his belief that, given the diversity of the United States and the medical care system’s capacity to generate revenues, a tightly regulated system makes the most policy sense. Vladeck, who holds a doctorate in political science from the University of Michigan and who has taught at Columbia University, concedes that—in this era of deregulation—he remains an unrepentant regulator. Vladeck, who became president of the United Hospital Fund, New York City, a year ago, was formerly an assistant vice-president of The Robert Wood Johnson Foundation. He is perhaps best known in the health policy community for managing the implementation of New Jersey’s hospital payment system based on diagnosis-related groups. Vladeck also is the author of a book published in 1980 by Basic Books entitled, Unloving Care: The Nursing Home Tragedy. The book is a critical exploration of the policy issues surrounding the nursing home industry.
In his Report to the Congress on Hospital Prospective Payment for Medicare which provided the basis for legislative enactment of the new Medicare prospective payment system, then-Health and Human Services Secretary Richard S. Schweiker made a special point of noting that: “An examination of Medicare records shows that payments for treating a heart attack average $1,500 at one hospital and $9,000 at another hospital with no apparent difference in quality. Likewise Medicare payments for hip replacements can vary from $2,100 to $8,200 and payments for cataract removal vary from $450 to $2,800. If Medicare is to become a prudent buyer of hospital services, it should pay the same price for comparable services.”

There is, one might argue, something of a logical leap from the Secretary's evidence and premises to uniform, all-inclusive national diagnosis-related group (DRG) rates, but no one really expects public policymaking processes to conform to all the canons of classical deductive logic. In fact, knowledge about variations such as those reported by Secretary Schweiker has played an important role in both the design and the defense of regulatory initiatives addressed at health care costs for several years, and is likely to play a still larger role in the future. The work of John Wennberg and his colleagues, beginning with the seminal article in Science in 1973, and perhaps more importantly—given the particular audience—comparative data on utilization patterns and length-of-stay variations developed by the Professional Standards Review Organization (PSRO) program after 1975, have substantially colored the perceptions and attitudes of regulatory officials.

This article seeks to describe that influence from the perspective of a former regulator. I will then examine some of the dead ends to which existing initiatives can lead. This article concludes by suggesting that while variations may be a big part of the problem, they are of only limited help in designing the solution.

The Regulatory Rationale

If one is concerned—for whatever reasons—with the rate of increase in hospital costs, the simplest and most direct approach is to cap the rate of increase at some arbitrary level, and apply that cap equally to all institutions. There is much to be said for such a method. It deals directly with the problem to be solved, and certainly minimizes administrative costs. But that kind of caricature of price controls has never been effectively sustained for long in hospital regulation. At the level of political rhetoric, not to mention sound policy analysis, it suffers from two fatal flaws.

First, existing patterns of hospital costs and prices already contain so much unexplained variation, regardless of how many factors are controlled for, that prospectively treating all institutions alike appears to
treat the efficient and inefficient equally. Worse, because of the effects of compounding, a simple uniform increase will reward over time the historically fat at the expense of the lean. That argument was used to particularly telling effect by opponents of the Carter administration’s cost-containment proposals in 1978 and 1979.

Thus, in the regulators’ jargon, it is necessary to start from a “clean base.” When, as is generally the case, hospital costs, using whatever denominator (days, discharges, DRGs, skull series) and controlling for all potentially explanatory factors (location, hospital size, teaching status) still exhibit a substantial range of variation, regulators can defend the presumption that the higher end of the range is unreasonable, barring some appropriate justification. The burden of proof is thus shifted to the high-cost institution.

The presumptive unreasonableness of statistically unexplainable high costs is not only analytically defensible. More important, as a practical matter, it is legally defensible. Reasonableness is an analytic and policy virtue, but it is also an essential requirement for any regulatory scheme that must meet basic constitutional tests. In an industry with such enormous preexisting cost variations, it is unreasonable, without substantial additional justification, to freeze that pattern into place. It is not unreasonable, in the legal sense, to statistically identify the upper end of that range and deem it excessive prima facie, provided that those in the upper end are given the opportunity of demonstrating that their higher costs are attributable to special circumstances.

Indeed, in the most stringent regulatory environments, such as New York’s, one can even defend the establishment of a standard of performance at the statistical median. Any costs in excess of that median, plus perhaps a narrow corridor to account for ordinary random variation, then appear excessive. After all—the regulators’ argument goes—half the hospitals can do as well, or better. Hospitals that cannot perform at least as well as the median better have a pretty good reason for their higher costs. Every institution has its own historical quirks and idiosyncrasies which cause its reported costs to vary from the norm, but occasionally there are special circumstances that do, by any measure, justify exceptions. A partially depreciated roof springs a leak; a visible incident creates the need for increased security; an inner-city institution has trouble discharging chronically ill patients to an inhospitable environment—every experienced regulator can cite multiple examples, and many more examples of less justifiable circumstances.

To a considerable extent, regulators have been able to sustain this line of argument because of its connection to the second major concern often expressed about more arbitrary price controls—their potentially deleterious effect on the quality of services. The quality of health care services is so self-evidently important and so widely perceived as a vital public
interest, that any regulatory intervention must be capable of being defended in terms of its effects on quality if it is to survive. Fortunately, from the regulators’ perspective, nobody is really capable of measuring the quality of hospital services very well, and to the extent that people believe they can at least distinguish among hospitals on qualitative criteria, there appears to be no systematic correlation between quality and cost. Just as the costs of hospital care vary dramatically from one institution to another within a given market, so too—it is believed—does the quality of service. But only rarely does the variation along these two dimensions coincide.

As far as anyone can tell at this point, there is some relationship between cost and quality in hospital services, but it is complicated, multidimensional, and sporadic. If the highest-cost hospitals were always the best, regulators would, indeed, have a serious intellectual and political problem. But so long as some reputedly very good hospitals are cheaper than some with no better reputations, justifying rate controls becomes considerably easier.

Regulators, in fact, lean very heavily on what might, with all appropriate respect, be called the “Johns Hopkins effect.” The Johns Hopkins Hospital has an unchallenged, and deserved, reputation for superior care. It is very expensive. But it is not the most expensive hospital in greater Baltimore for many of the things it does. To any hospital which is more expensive, after all the appropriate arcane adjustments, in treating any given diagnosis or performing any given procedure, the regulator can (and generally will) always say: “If Hopkins can do it for x dollars, why can’t you?” And without any embarrassment or hypocrisy, the same regulator can turn to Johns Hopkins and say: “If you can treat condition for costs no greater than the adjusted median for teaching hospitals, why can’t you do as well with condition y?” Of course, if interinstitutional variation were more systematic or more predictable, such arguments could not be made, and the “Johns Hopkins effect” would be nullified. But it hardly ever is.

Indeed, the most striking thing about interinstitutional variation in costs, quality, and treatment patterns is how much of it there is, how unsystematic it appears to be, and how much of it remains unexplained by even the longest and most complicated regression equations. In its simplest form, most hospital rate regulation to date has focused on squeezing the high end of that variation out of the system.

### The Road To DRGs

The extreme example of such squeezing, of course, is provided by national diagnosis-related group (DRG) rates. And as the extreme case, they offer a useful paradigm for exploring some of the limitations of the
prevailing regulatory logic.

Essentially, the logic for uniform national DRG rates holds that if half of all the cases in the nation in a given DRG can be treated for x dollars or fewer, then it is not unreasonable or unfair to expect all hospitals to do as well. Moreover, it would be imprudent, if not downright stupid, for a public monopsonist to expend any more tax dollars. To the argument that hospitals that lose money on all DRGs will be imperiled, the federal bureaucrat can respond, accurately, that as a matter of empirical fact, few hospitals will perform poorly on all DRGs; most hospitals will do well on some, less well on others. But that argument, when examined, vitiates the case for too much uniformity in rate setting.

Putting aside, for the moment, the very severe shortcomings in Medicare’s DRG cost-allocation methods, the fact that the variation in costs within DRGs within hospitals mirrors the variation within DRGs across hospitals suggests that what is really going on is a lot of variation that, for all intents and purposes, is pretty random. One can strike a mean and draw a lopsided bell curve of DRG costs, but that curve is rarely very peaked. And indeed, one can generally strike a mean and draw a bell-shaped curve around almost anything if one’s sample size is big enough; in Statistics I, that phenomenon is described as the Law of Large Numbers. (The best, and most defensible way to reduce the variance around mean costs or length-of-stay within any DRG is to broadly define outliers, and leave as “inliers” only a very homogeneous group. That approach has been explicitly rejected by the federal government in its DRG system.)

As a classification system, DRGs are often criticized for the amount of variance, or heterogeneity, remaining within many of the 468 groups, but that criticism really misses the point. It is not that the classes are too imprecise; the problem is that the world out there is extraordinarily imprecise and pervasively characterized by significant variations. Indeed, one of the best defenses of DRGs as the basis of a payment system is that they leave rather less unexplained variance than do most other plausible categorization schemes.

Thus, in the development of the New Jersey DRG system, we believed very strongly that DRGs made an excellent unit of payment, but tried hard to resist pressures towards homogeneity in pricing in excess of what the underlying reality seemed to make defensible. We based each hospital’s rates to a considerable degree on its own costs, kept about a third of all hospital costs outside DRG pricing, and created many (in hindsight, too many) outliers. And New Jersey is, medically speaking, a pretty homogeneous place—at least in comparison to all of the United States.

To put this argument another way, the phenomenon of cost variations across hospitals—when costs are expressed in DRGs or any other unit—makes it hard to justify freezing existing relative prices into place under a
regulatory system. But the very extent and nature of that variation also argues against too much uniformity, or at least too much uniformity arrived at too quickly. No one knows why there is so much variation, but it is certainly there, and needs to be reckoned with.

### The Problem With Prices

Most of what is commonly thought of as the literature on variations in medical practice patterns and services utilization has, of course, little directly to do with variations in costs or prices. Most of it speaks, instead, to variations in use rates—in surgical procedure rates, or ancillary service consumption, or total hospital days per ‘unit of population. This is not a different kind of phenomenon from price variation, but an additional one. And again there is distressingly little covariance. Communities with high unit costs are sometimes also those with high usage rates, but sometimes are those with low usage rates, though not often enough. For Medicare recipients, Miami has high use rates and high unit costs; Tacoma has low unit costs and low usage rates. But the rest of the nation’s metropolitan areas are all over the lot.

The important difference between price variation and use variations is their denominators: service units for price, population for use rates. The important policy issue is that both price and use are critical elements of the total cost equation. Expenditures, after all, equal quantity times price. Controlling one does very little good if the other balloons in its place, as it probably will. Control prices on a per diem basis and length-of-stay will increase; control them on a per case basis and admissions will probably increase. On the other hand, if you effectively control utilization, unit costs—if not prices—will increase, at least in the short run. Health maintenance organizations (HMOs) can succeed, in areas where their market share is relatively small, precisely because any given hospital can trade off price reductions for guaranteed volume.

The Medicare DRG system, like most state regulatory systems, controls prices only, and that is, of course, its most serious weakness. In states with relatively mature price regulation, regulators and budget officials tend to be preoccupied with volume, either directly, in the form of utilization controls, or indirectly, in the form of capital expenditure controls. Medicare officials will soon be similarly concerned.

The need to get simultaneously at both volume and price is clearly why everyone is so enamored of capitation. In any given community, capitation limits the payer’s risk that controlling one half of the equation will lead to inflation in the other, and makes it the capitated plan’s problem. In the absence of coercive measures, subtle or overt, it is, fortunately or unfortunately, hard to create effective capitated plans willing to accept such risks over time, and harder still to get lots of people—or
at least lots of sick ones—to enroll in such plans, but that’s a whole other set of problems not directly relevant here. What is directly relevant is the problem that, from the perspective of national policy, widespread capitation simply would raise the variations issue from the institutional or community level to the regional one. Schweiker’s diagnostic examples would be still more troublesome politically if a similar range of variation developed for capitation rates across congressional districts. At the same time, capitation doesn’t solve the problem of intraregional variations, either. It just internalizes the problem in some hypothetical, yet to be developed, “managed system.”

The Shape Of Future Policy

If every community in the United States had a cardiac catheterization rate identical to that of Olmsted County, Minnesota, a lot fewer people would have their coronary arteries catheterized each year, and a lot of money would presumably be saved. The catheterization rate in a community served by the Mayo Clinic provides a good example for a regulator or payer in a community with a much higher rate, adjusted for age, sex, race, and other factors. He can sincerely tell the public that a lower rate is unlikely to hurt them very much, and he can probably choose to pay for only, say 90 percent of his community’s rate. He has, in short, an awful good club with which to pound providers. But what else it may tell us is not entirely obvious. Some of us, after all, are prepared to pay a lot of money, not to mention making other sacrifices, in order not to live in Olmsted County.

The fundamental problem, again, with seeking to discern a policy direction from our knowledge of variations is not that variation exists, but that there is so much of it, too much for purposes of intellectually satisfying policy development.

Three somewhat broader conclusions emerge. First, and most obvious, is the extent to which the actual practice of medicine remains so much more art than science, and an art that is dominated by different schools in different places. One can seek to control the costs for art as much as the costs for science, but one probably should go about it in a somewhat different way. A too systematic concern for uniformity, regularity, and rigid standards is probably, at best, a waste of energy.

Second, the growing evidence on interarea variations raises increasing questions about the wisdom or desirability of having national policy at any but the most global level. The United States remains a very heterogeneous nation, and it may well be simply unrealistic to expect great homogeneity in health care behaviors or health care prices than we have long accepted in education or public transportation or, for that matter, Medicaid. One can still make a very strong case for a uniform national
level of minimal insurance coverage and health care availability, but the case for very much more uniformity is much harder. The argument is often suggested that, under the status quo, taxpayers in Tacoma are subsidizing Medicare beneficiaries and (especially) providers in Miami, but Miami taxpayers subsidize Tacoma electricity users, loggers, and Boeing employees. In a continental federal system, what else should one expect?

Finally, thinking through the implications of variations data for cost-containment policy eventually leads us back to where we started. The current level of health care expenditures in the United States is almost certainly higher than it needs to be—the data on variations clearly demonstrates that—but the excesses embodied in the current level are not the real policy problem. The real policy problem is the rate of increase in those costs. And if one wishes to control the rate of increase in costs, the way to do that is to control the rate of increase.

What the data on variations do appear to tell us is that a series of limits, however arbitrary in appearance, are unlikely to produce results, in the short to medium term, any more capricious or random than what we now have. Fairness demands provision for making exception to arbitrary caps when circumstances so justify. It demands “cleaning the base” to avoid freezing inordinately high costs into place. It does not demand uniformity. The instinct of researchers and analysts—and increasingly, if unfortunately, of policymakers as well—is to react to randomness and confusion by seeking to simplify, to glorify the mean (or median), to seek out uniformities when heterogeneity more accurately characterizes the world. Doing so may make for good science and good analysis. It does not necessarily make for good policy.

NOTES