DEALING WITH MEDICAL PRACTICE VARIATIONS: A PROPOSAL FOR ACTION

by John E. Wennberg

Prologue: *Without much attention from the profession and virtually no public fanfare, John Wennberg has been tracking the phenomenon of variations in the use of medical care for more than a decade. Wennberg, who ranks among the leaders of the nation's tiny cadre of medical care epidemiologists, has been driven by the notion that practice variations were important to identify and understand because they suggest a misuse of care Wennberg trained at Johns Hopkins University in internal medicine and also holds a master's degree in public health from that institution. Currently, he is a professor in the Department of Community and Family Medicine at Dartmouth Medical School. From 1974 to 1979, he served on the faculty of the Harvard Medical School, attached to the Department of Preventive and Social Medicine. During his research pursuits, Wennberg has uncovered systematic and persistent differences in the standardized rates of use for common surgical procedures and other medical services in the United States. But this phenomenon is not limited to any particular health care delivery system. Such variations have been found wherever Wennberg and his colleagues have looked, be it New England, Iowa, the United Kingdom, or Norway, as well as among the political subdivisions within these areas. Wennberg believes that the principal reason for the dramatic variations in use of medical care can be found in what he characterizes as the "practice style factor." Wennberg's conclusion has particular relevance to Medicare's new hospital payment approach bused on diagnosis-related groups (DRGs). Built into this approach to cost containment is an assumption that the mix of hospitalized patients is largely unaffected by physician practice styles. But, as he has found, professional discretion is a very important variable and must be taken into consideration in dealing with the health cost conundrum.*

Most people view the medical care they receive as a necessity provided by doctors who adhere to scientific norms based on previously tested and proven treatments. When the contents of the medical care "black box" are examined more closely, however, the type of medical service provided is often found to be as strongly influenced by subjective factors related to the attitudes of individual physicians as by science. These subjective considerations, which I call collectively the "practice style factor," can play a decisive role in determining what specific services are provided a given patient as well as whether treatment occurs in the ambulatory or the inpatient setting. As a consequence, this style factor has profound implications for the patient and the payer of care.

For example, the practice style factor affects whether patients with menopausal symptoms, with hypertrophy of the tonsil, with hyperplasia of the prostate, with mild angina, or with a host of other ailments receive conservative treatments in an ambulatory setting or undergo a surgical operation in a hospital. It also affects whether patients with relatively minor medical conditions such as bronchitis or gastro-enteritis, or who need minor surgical procedures such as cystoscopy, teeth extractions, sterilization, or breast biopsy receive their care in a hospital or elsewhere. The practice style that favors inpatient treatment greatly affects the demand for hospital care and has serious implications for efforts to constrain costs.

These implications become clear when one recognizes that, within a region or state, different opinions held by physicians concerning the need for hospitalization–as measured by per capita admission rates–are the most important determinant of variations in per capita costs for the treatment of specific diseases.[1] The different opinions of doctors over the need to hospitalize are much more influential in established total costs than differences in cost per case or the length of an inpatient stay.

Some of the differences in opinion arise because the necessary scientific information on outcomes is missing: controversies about alternative therapies cannot be resolved through appeal to existing evidence. To resolve the differences in opinion–and to learn whether high or low rates of admission reflect appropriate care–more scientific information must be obtained.

For other conditions, the practice style factor appears unrelated to scientific controversies. Physicians in some hospital markets practice medicine in ways that have extremely adverse implications for the cost of care, motivated perhaps by reasons of their own or their patients' convenience, or because of individualistic interpretations of the requirements for "defensive medicine." Whatever the reason, it certainly is not because of adherence to medical standards based on clinical outcome criteria or even on statistical norms based on average performance. In

some markets, a substantial proportion of hospitalizations are for cases that in other markets are usually treated outside the hospital. If more conservative, ambulatory-oriented practice styles were substituted–and if hospital administrators and trustees translated the decreased demand for hospital resources into a reduction in the capacity of the hospital industry–then substantial cost savings and improvements in quality could be realized without fear that needed services were being withheld.[2]

In this article, I propose a plan for dealing with the health care outcome and cost-containment implications of the medical practice variation phenomenon. My goal is not to obliterate all variation from the practice of medicine. Obviously, physicians must have freedom to apply the skills of their profession as they and their patients see fit; after all, medicine remains as much art as science. Moreover, any enterprise as large as medical care will produce variations in approach. My targets are variations that are both substantial and reflective of supply factors rather than scientific knowledge and the values, needs, or wants of patients.

My plan has three parts. The first calls for a closer monitoring of medical practice in local hospital markets, using epidemiologic techniques to obtain population-based measures of resource allocation, service use, and outcomes of health care. This information should be made available on a continuous basis to interested parties. Second, I recommend that the medical community and qualified researchers address unanswered questions concerning the effectiveness of many common therapeutic interventions. The overriding questions in this regard are whether such interventions have beneficial outcomes and are relatively safe.

Third, I recommend that the medical community make greater efforts to deal with the cost-containment problem by reducing the use of hospitals for marginally indicated conditions, as may be determined from the monitoring of medical practice called for above. The challenge would be to translate these reductions in inpatient demand into reductions in the capacity of the hospital industry as a step toward moderating the growth of per capita health costs.

In advancing this plan, I draw on my experience with monitoring the performance of the medical care systems in New England over the past decade and, more recently, in Iowa. In these areas, I have worked closely with doctors and state medical societies to feed back to physicians the information I found. The positive physician response to this information suggests that doctors and their professional organizations in other areas can be expected to assume leadership roles in projects that deal with the cost and medical outcome implications of the variation phenomenon. But the feasibility of the plan will depend ultimately upon broad-based support from the private and public sector, including government. I offer some specific suggestions on the nature of that involvement.

Examining The Medical Care "Black Box"

Before discussing the plan, I want to review the evidence that argues for attaching importance to supply factors in determining the demand for hospitalization. I also want to examine the variation phenomenon in greater detail, particularly its implications for cost containment.

Why is it that the norms of medical practice can be so loose or ambiguous as to allow a wide range of professional discretion? The answer is seen in a review of the medical literature on the degree of professional consensus concerning the value of specific treatments and in the critical examination of the scientific strengths and weaknesses of the studies which support a particular viewpoint.

The procedures exhibiting the most variation are often for conditions that are part of the aging process. The controversies arise because for such conditions the natural history of the untreated or conservatively treated case is often poorly understood, and well-designed clinical trials are notably absent. Examples include hysterectomy for noncancerous conditions, prostatectomy for benign hyperplasia of the prostate, tonsillectomy for hypertrophy of the tonsil, and coronary bypass surgery for mild angina. Well-defined scientific norms simply do not exist to limit the practice options physicians select to treat these maladies. As a consequence, the opinions of individual doctors can vary substantially, based upon subjective experience. Because many of the conditions are associated with the aging process, the number of candidates that could qualify for operative intervention is sometimes upwardly limited only by the size of the population.

For example, I have observed that in Maine, by the time women reach seventy years of age in one hospital market the likelihood they have undergone a hysterectomy is 20 percent while in another market is is 70 percent. In Iowa, the chances that male residents who reach age eighty five have undergone prostatectomy range from a low of 15 percent to a high of more than 60 percent in different hospital markets. In Vermont the probability that resident children will undergo a tonsillectomy has ranged from a low of 8 percent in one hospital market to a high of nearly 70 percent in another.

By contrast, the low-variation procedures derive from quite specific conditions for which there is a professional consensus on the preferred place or style of treatment. Prime examples are surgical repair of inguinal hernia and hospitalizations for hip fractures. For these conditions, practice style, at least in the United States, only plays a small part in affecting demand. As a general rule, diagnosis is not difficult. If the patient seeks medical care, variations in clinical judgments are constrained by a consensus. For inguinal hernia, the treatment is an operation. For hip fraction, virtually all patients are hospitalized.[3] The rates show little variation

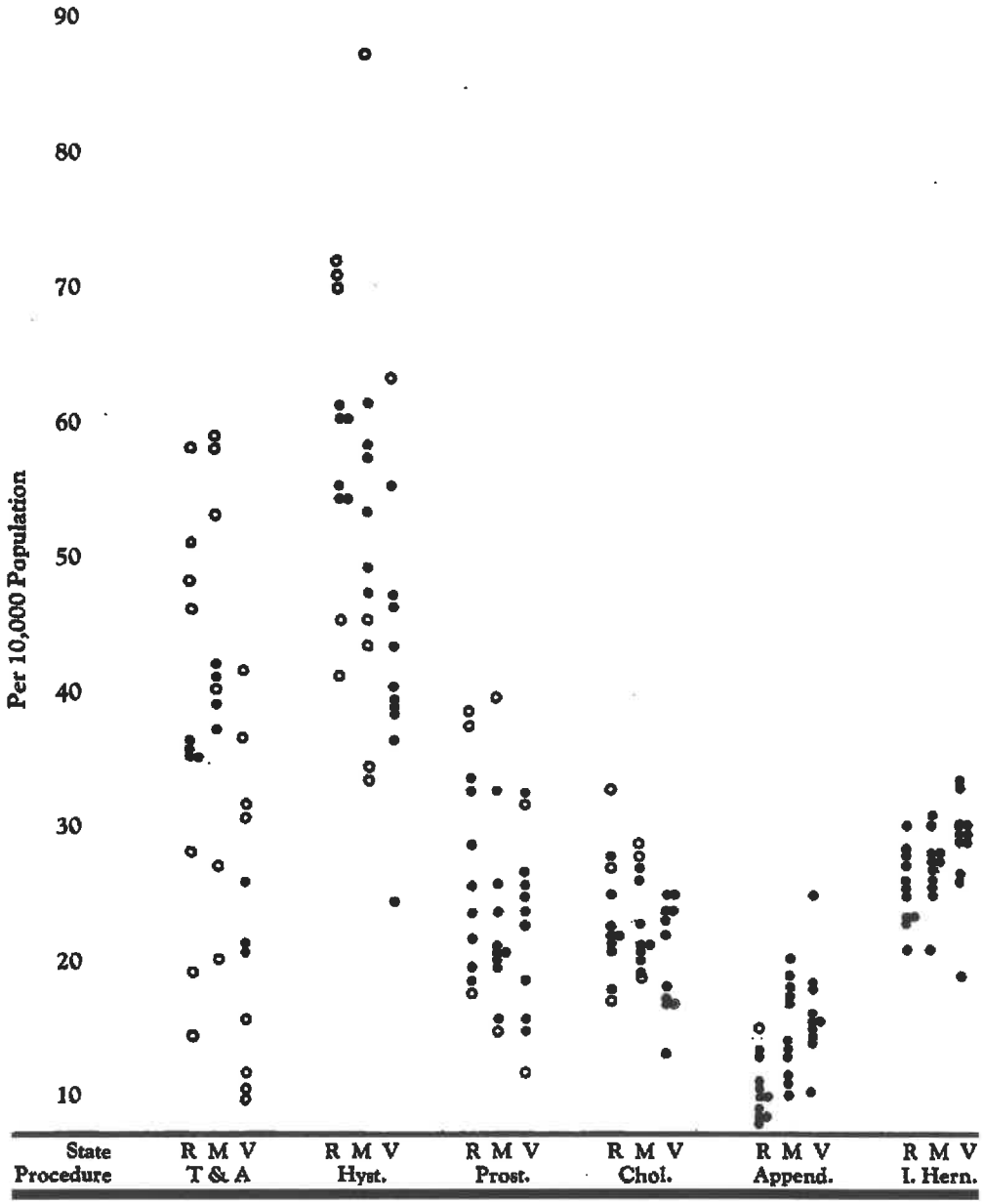
between hospital markets. The pattern of variation for common procedures in three New England states is illustrated in Exhibit 1.

I developed the practice style theory after it became clear to me that the variation phenomenon could not be explained adequately by traditional theories. For example, consumer or population factors do not explain much of the difference in utilization rates among local hospital markets. Household interview studies in Vermont and Manitoba, Canada compared the characteristics of residents living in high- and low-rate market areas. These studies failed to show correlations between service use rates and illness rates, insurance coverage, access to service, and other demand-related attributes of patients or populations. The variations also persist after adjustment for age, which tends to account for most illness-related differences in populations. While some variation in the use rate of specific procedures is explained by the per capita supply of facilities and physicians, most remains unexplained. For example, Noralou Roos has shown that per capita number of beds and gynecologists are virtually the same in hospital markets with low and with high per capita hysterectomy rates.[4]

The variation phenomenon appears to be a worldwide phenomenon, not explained by incentives associated with fee-for-service medicine. The pattern of variation for common procedures is similar among fee-for-service hospital markets in Iowa, Maine, Massachusetts, Rhode Island, and Vermont; among health maintenance organizations in the United States; and among the health care regions in Canada, England, and Norway, even though obvious differences exist in the supply of surgeons, the organization and financing of services, and in the cultural and demographic characteristics of hospital market area residents. There is one common factor: physicians in each of these areas of the world read the same medical literature, participate in the same scientific traditions, and share the same scientific uncertainties concerning the value of certain procedures.

The most direct evidence for the importance of practice styles in influencing utilization rates comes from natural experiments in which practice styles change following the feedback of information to physicians on the rates in their own and neighboring market areas. Changes have been documented for hysterectomy rates in Saskatchewan, Canada and Maine; for tonsillectomy rates in Vermont and Maine; and for lens extractions in Norway. The evidence indicates that the changes occurred primarily because physicians took actions to modify their clinical policies.[5] In one example, a letter from hospital officials speaks directly to the importance of admission policies in influencing the demand for services and documents the effect of feedback. The letter reads: "We are following up after an ongoing one-year audit (of) hysterectomies . . . (concerning) the high numbers of hysterectomies. . . (The past experience) in no way reflects

Exhibit 1
Age-Adjusted Rate Of Procedure For Six Common Surgical Procedures In Rhode Island, Maine, and Vermont (1975)



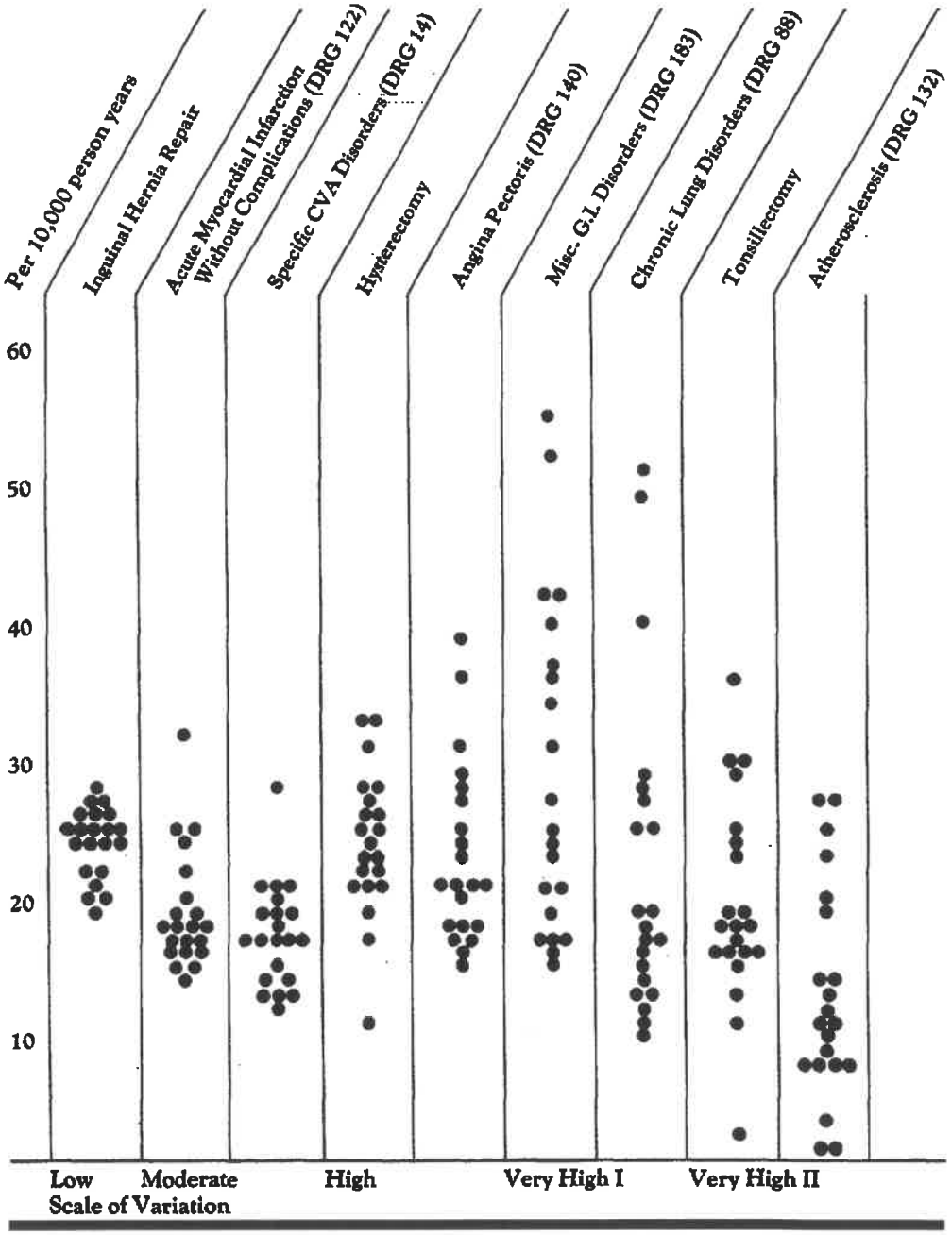
Note: Rates of surgical procedures vary greatly among hospital areas. The rates shown are for the six most common surgical procedures for the repair or removal of an organ in the eleven most populous hospital areas of Maine, Rhode Island, and Vermont (1975). The rate of tonsillectomy varies about sixfold among the thirty-three areas; the rates of hysterectomy and prostatectomy vary about fourfold. Moreover, many of the extreme rates for these procedures differ from the average rate for the state by an amount that is statistically significant (open circles). There is much disagreement among physicians on the value of the high-variation procedures. Similar patterns of variation for these procedures have been observed in Iowa, England, and Norway. R = Rhode Island; M = Maine; V = Vermont.

the (current) actual numbers. The department of OB/GYN has set its own goals of between 220 and 240. This past year, 229 hysterectomies were completed during that period. Also, we have met our own criteria regarding the number of hysterectomies showing normal pathology at 20 percent to 25 percent. During this one-year period, a percentage of 24.9 was reached." In the year prior to feedback, the market area served by this hospital had a rate of 118 procedures per 10,000 women which was more than double the state average. In 1981, the year the letter discusses, the rate was 58, less than half the previous rate but 25 percent higher than the 1981 state average.

The value of the practice style theory is further illustrated by its power to provide a reasonable interpretation for the relative variation observed for rates of hospitalization. Exhibit 2 and Exhibit 3 show the relative variation of rates for cases classified by cause of admission, using a modification of the diagnosis-related group (DRG) case-mix classification system. Clinicians, in reviewing the information, will recognize an association between the order of the listing of causes of admission, ranked by the measure of variation in rate of admission, and the degree of discretion that physicians can exercise in the decision to hospitalize or not. At the lower end of the variation scale are admissions for hip fracture and for myocardial infarction. For these conditions, there is little choice and under current standards for care in the United States, patients must be hospitalized. But in the high range of variation–those causes of admission that are more variable than hysterectomy–the situation is not so clearcut. For example, many cases of bronchitis, or fracture of the forearm, can be and often are treated in the ambulatory setting; it is quite reasonable to infer that the fivefold range in variation for these causes of admission seen in Exhibit 2 reflects differences in local practice styles in how this decision is made. For the examples of causes of admission that rank at the extreme high end of the variation scale, the data speak for themselves on the governing importance of professional decision making in determining the use rates of hospital resources. Hospitalizations for tooth extractions and restorations, and for pediatric admissions for gastro-intestinal diseases, show a twentyfold range of variation.

Exhibits 2 and 3 illustrate another important feature of the variation phenomenon: high-variation profiles are the rule not the exception. When the "black box" is examined, most of the individual contents of medical practice–as defined by conventional classification systems–are more variable than hysterectomy. For example, when all medical, surgical, and pediatric hospitalizations are examined, less than 13 percent of cases are for causes that show less variation than hysterectomy; 12 percent are more variable than tonsillectomy. Most surgical as well as diagnostic procedures are also high variation. So much for the idea that most medical services are undifferentiated necessities, dispensed according to scien-

Exhibit 2
Age-Adjusted Incidence Of Hospitalization (1980–82)
For Selected Medical DRGs And Three Common Surgical Procedures
For Maine Hospital Markets



Note: The DRGs with similar statewide rates were selected to demonstrate the spectrum of variation in the incidence of DRG-specific hospitalizations among Maine hospital markets. Each circle represents a hospital market area. The graph is limited to markets with 45,000 person years or greater. The incidence of hospitalization for most DRGs is more variable than for hysterectomy.

Exhibit 3
Medical And Surgical Causes of Admissions Ranked In Ascending Order Of Variation In Incidence Of Hospitalization (1980–1982)

Medical Causes of Admission

Low Variation
None

Moderare Variation
Acute myocardial infarction
Gastro-intestinal hemorrhage
Specific cerebrovascular disorders

High Variation
Nutrional and metabolic diseases
Syncope and collapse
Respiratory neoplasms
Cellulitis
Urinary tract stones
Cardiac arrhythmias
Miscellaneous injuries to extremities
Angina pectoris
Toxic effects of drugs
Psychosis
Heart failure and shock
Seizures and headaches
Adult simple pneumonias
Respiratory signs and symptoms
Depressive neurosis
Medical back problems
Digestive malignancy
G.I. obstruction
Adult gastro-enteritis
Peripheral vascular disorders
Red blood cell disorders
Adult diabetes
Circulatory disorders exc. AMI, with card cath.

Very High Variation
Deep vein thrombophlebitis
Adult bronchitis and asthma
Organic mental syndromes
Chest pain
Transcient ischemic attacks
Kidney and urinary tract infections
Acute adjustment reaction
Minor skin disorders
Trauma to skin, subcut. tiss. and breast
Chronic obstructive lung disease
Hypertension
Adult otitis media and URI
Peptic ulcer
Disorders of the biliary tract

Medical Causes of Admission (cont.)

Pediatric gastro-enteritis
Pediatric bronchitis and asthma
Atherosclerosis
Pediatric otitis media and URI
Pediatric pneumonia
Chemotherapy

Surgical Causes of Admission

Low Variation
Inguinal and femoral hernia repair
Hip repair except joint replacement

Moderare Variation
Appendicitis with appendectomy
Major small and large bowel surgery
Gall bladder disease with cholecystectomy
Adult hernia repairs except inguinal and femoral

High Variation
Hysterectomy
Major cardiovascular operations
Pediatric hernia operations
Hand operations except ganglion
Foot operations
Lens operations
Major joint operations
Stomach, esophageal, and duodenal operations
Anal operations
Female reproductive system reconstructive operations
Back and neck operations
Soft tissue operations

Very High Variation
Knee operations
Transurethral operations
Uterus and andenexa operations
Extra-ocular operations
Misc. ear, nose, and throat operations
Breast biopsy and local excision for nonmalignancy
D & C. conization except for malignancy
T & A operations except for tonsillectomy
Tonsillectomy
Female laparoscopic operations except for sterilization
Dental extractions and restorations
Laparoscopic tubal interruptions
Tubal interruption for nonmalignancy

[a]Causes of hospitalizations are taken from Diagnostic-Related Disease Classification system but cases have been grouped without regard to presence or absence of significant complication. Obstetrical and neo-natal causes of hospitalization are excluded. Ranking is according to the Systematic Component of Variation. Variations are measured across thirty hospital markets. The exhibit lists individually only those with more than 1,500 cases. More than 50 percent of hospitalizations are represented in the exhibit. Classes of variation are defined such that the variation associated with the first entry in a class is significantly more variable than the first entry in the previous class. For additional information see K. McPherson, J.E. Wennberg, O.B. Hovind, and P. Clifford, *The New England Journal of Medicine* 307(1982):1310-4.

tific norms.

To understand the cost implications of the practice style factor, return to the case of hysterectomy. Over the last decade, about 2,500 more hysterectomies were performed in one Maine market area, with about 35,000 women in the eligible age range, than would have been performed if the practice style of physicians located in a low-rate neighboring market had applied. At the current market value of approximately $4,000 per hysterectomy, this translates into an excess of $10 million. Yet the price of hysterectomy in this high-rate market is below average. Under DRG prospective pricing plans and preferred provider strategies, the hospitals in the high-rate market will be rewarded for their "efficiency." Yet the data show that the most important determinant of variations in per capita costs, or the "bottom line" for payers, are physicians' decisions to admit patients to the hospital or to employ a specific treatment, not the decisions they or other health care providers may make that affect the efficiency of medical care as reflected in the unit price of service or the length of a hospital stay.

Since more than 85 percent of hospitalizations classified under the DRG system appear to have greater variation in per capita use rates among hospital market areas than hysterectomy, the above example is a conservative demonstration of the problems that attend unit pricing approaches to cost containment. Indeed, through the incentives they create for reducing lengths-of-stay in hospitals–thus freeing beds to allocate to new patients–cost-containment programs that focus only on the reduction of unit price may add to rather than reduce the overall costs. If hospitals that stand to lose money under the DRG system are able to improve their financial status by increasing the volume of services, then the inevitable result will be an acceleration in the rate of increase in per capita expenditures.

The First Step: Monitoring Performance In Hospital Markets

Hospital markets are thus characterized by highly variable rates of use for most specific medical treatments, diagnostic tests, and surgical procedures, and by widely different resource use rates. The actions that are needed pertain to the clinical management and resource allocation decisions in specific hospital markets. The first step is to monitor and distribute information on the per capita performance in local hospital markets so that decision making can be modified when appropriate.

What are the essential features of the monitoring I propose? The necessary data are contained in health insurance records such as Medicare, Medicaid, and Blue Shield claims systems and hospital discharge abstracts similar to those used in the DRG program. Population counts and information on hospital resources, including annual budgets, numbers of

facilities and personnel also are needed. For outcome reports, information on survival must be joined to discharge data and to claims data to establish the link between use of medical care, diagnoses, and outcome. Sources for this information exist in many parts of the country and, for the Medicare program, nationwide.

The data should first be used to determine the geographic origin of patients who seek care at specific hospitals. The individual communities of a county or state are then arranged into hospital market areas such that most hospitalizations of local residents occur within area hospitals (and are thus initiated by physicians practicing within the area). Following this strategy, my colleagues and I have defined some 200 hospital markets in the six New England states and over 100 in the state of Iowa. The way the markets are organized assures a close association between the medical care experience of the local population and decisions made by health planners, regulators, local administrators, hospital trustees, clinicians, and, potentially, business coalitions. Since information on resource allocation and service use rates is available from all relevant places where care is given (whether in- or out-of-area), the per capita rates are truly population-based and thus may be validly compared.

What do the reports look like? There are three kinds of reports. One series describes the status of resource allocations to specific communities: the number of hospital beds, expenditures, and hospital personnel or the number of physicians invested, per capita in the health care of the local communities. Exhibit 4 is an example for Boston, Massachusetts and New Haven, Connecticut. Comparisons such as this should be very useful in planning decisions concerning capital expansion projects and in setting hospital budgets under prospective reimbursement plans.

In reviewing the reports, it is important for the reader to understand that virtually all of the hospitalization experience of the resident population is accounted for even if it takes place at hospitals located in other areas. The reports can be used to project the per capita consequences of specific planning or regulatory decisions. They can also be used in cost-containment strategies to reduce expenditures in high-cost markets by cutting or stabilizing the size of the local hospital industry as indicated by its contribution to the total numbers of personnel and beds per capita. Variations in these indicators are strongly correlated with per capita expenditures; with this information, hospital administrators and trustees can make a direct connection between plant size and employment complements in their specific hospitals and the variations in the total per capita costs.

The reports inevitably raise issues concerning the relationship between the quality of care and the level of resource investment, particularly if the comparisons are between markets with a high proportion of patients who are treated in a university teaching hospital. In Exhibit 4, most

Exhibit 4
The Quantity Of Hospital Resources Expended On The Populations Of New Haven, Connecticut And Boston, Massachusetts By Hospitals Providing Resources (1978)

Hospital	Percent of admissions from the local pop.	Beds allocated to local population	Market share	Beds	Per Capita Rates[b] Expend.[a]	Personnel
New Haven, Connecticut (pop. est. 372,900)						
Yale-New Haven Univ. Hosp.	68.3	541.6	54.8	1.5	124	5.5
St. Raphael	86.4	416.6	38.1	1.1	82	3.5
Out-of-area hospital		65.0	7.1	0.1	9	0.5
All hospitals		1023.2	100.0	2.7	215	9.5
Boston, Massachusetts (pop. est. 732,400)						
Boston teaching hospital (N=7)	42.6	1828.0	59.0	2.5	322	13.1
Boston community hospital (N=11)	50.6	843.0	23.3	1.2	84	3.3
Out-of-area hospital		524.4	16.7	.7	42	1.8
All hospitals		3195.4	100.0	4.4	448	18.2

Note: The estimates for the resources allocated to the New Haven and the Boston populations are made by multiplying the amount of resources provided by each hospital by the percent of admissions that are from the local population (column 2). For example, 542 of the Yale-New Haven University Hospital's total complement of 793 beds are used by the residents of New Haven. The estimate for the total numbers of beds is obtained by summing column 3 which, it will be noted, includes beds from out-of-area hospitals that provided services to the population of New Haven. For comparative purposes, we are particularly interested in per capita rates. The exhibit shows these for beds, numbers of personnel, and inpatient expenditures. All rates are corrected for boundary crossing.

[a] For inpatient services.

[b] Beds and personnel per 1,000 population, expenditures per person.

resident hospitalizations are to well-known hospitals and it might be assumed that per capita costs in each market would be quite high. This is not the case. The New Haven market area ranks in the middle third of all market areas in Connecticut, largely because of its relatively low total numbers of beds and personnel per capita. Contrast this to the situation in Boston where the per capita expenditures are more than double: in New Haven, in 1978, the estimate was $215; in Boston it was $448. The beds allocated to the population of Boston number 4.5 per 1,000 while in New Haven they number only 2.7. The number of employees per 1,000 shows about a twofold variation.

The differences in resource use depicted in Exhibit 4 are apparent only when directly measured. They are not intuitively known by those on the scene. I have asked clinicians who have practiced in both Yale and Harvard teaching hospitals to estimate the per capita expenditures in each market. Their answers indicate they have no awareness of the magnitude of the difference; what is more surprising, many do not accurately guess which of the two markets is the more expensive. Nor can the differences be appreciated through the use of traditional indicators of performance, whose validity as measures of market consumption rates depend on the

Exhibit 5
Rates For Cystoscopies Among Maine Medicare Enrollees By Urology Market Area Of Residence (1976-1977)

Urology market area	Enrollees[a]	Number of examinations	Rate[b]	Ratio to state average	Percent of enrollees with one or more examination
Portland	43,192	1,641	3.8	1.33*	2.8
Bangor	29,814	857	2.9	1.00	1.8
Lewiston	16,397	328	2.0	.70*	1.5
Augusta	9,920	235	2.4	.83*	1.7
Waterville	12,886	201	1.5	.54*	1.2
Biddeford	8.212	315	3.8	1.34*	2.6
Rumford	3,895	232	5.9	2.08*	3.9
Presque Isle	6,361	143	2.9	.78*	1.6
Skowhegan	4,203	95	2.3	.79	1.6
Ellsworth	2,805	68	2.4	.85	1.5
Caribou	5,757	125	2.2	.76*	1.8
Calais	1,969	23	1.2	.41*	1.0
State	156,325	4,478	2.86	1.00	2.0

Note: The count of the number of cystoscopic examinations is made from the claims history files of the Medicare program obtained from the carrier, using the appropriate procedure codes to select the relevant records. Reimbursements (not shown) are also tabulated from the claims records. The population counts are for all Medicare enrollees who were in the Part B program in 1977. The percent with one or more cystoscopy is determined by counting enrollees with cystoscopic examinations, rather than number of services.
[a]Enrollee person-year.
[b]per 1,000 enrollees.

degree to which they correlate with the per capita market rates. Small area research indicates their virtual independence. For example, among the hospital markets of a state, the occupancy rates of local hospitals, their average lengths-of-stay, and such measures of efficiency as the number of patients treated per bed (properly weighted to measure each hospital's relative contribution to the total experience) show little relationship with per capita number of beds or patient days, inpatient expenditures, and reimbursements per capita.

A second series of reports is concerned with the utilization of specific services for surgical and diagnostic procedures and for causes of admission. Exhibit 5 gives an example for diagnostic procedures, showing the rate of use of cystoscopic examination among Medicare residents in twelve Maine markets defined for urology services (1976-1977). The exhibit is based on claims data from the Medicare Part B program and the Medicare enrollment file. Note that the cystoscopic rate in the Rumford market is more than double the rate for the state as a whole, while in the Waterville market it is only about 54 percent of the average. The range of variation for the volume component (the per capita use rate, given in the exhibit as the standardized procedure rate) varies by a factor of more than four

Exhibit 6
Admissions To Hospital For Medical Back Problems (DRG 243) And For Dental Extractions And Restorations (DRG 187) Number Af Cases Above (+) Or Below (–) Expected, Based On State Average Nine Most Populated Maine Hospital Markets (1986-1981)

Market areas	Back Problems			Dental Extractions		
	Admissions Observed –Expected[a]	Standard Rate	Reimburs. Observed –Expected (x$1000)	Admission Observed –Expected a	Standard Rate	Reimbure. Observed –Expected (x$1000)
Portland	–567.1*	.58	–1,048	– 149.8*	.49	–122
Bangor	–61.9	.91	–108	–81.9*	.48	–66
Lewiston	–283.9*	.59	–503	–108.2*	.29	–87
Augusta	+ 162.8*	1.32	+288	–42.0*	–60	–33
Waterville	+ 150.6*	1.37	+267	–55.1*	.43	–44
Biddeford	–74.0*	.81	–131	+88.6*	2.10	+71
Brunswick	–10.3	.96	–18	+115.9*	2.90	+93
Rockland	+2.7	1.01	+5	+60.0*	2.30	+48
Farmington	–74.0*	.66	–131	–37.7*	.24	–30
All Other	1755.1*	–1.27	+1,339	+172.1*	1.18	+140

Note: The input to the table is hospital discharge data, maintained by the Maine Health Information Center and population data from the 1980 census, DRG-specific reimbursement rates are estimated using charge data from the Maryland Hospital Cost Commission for 1980. Column 2 gives the actual number of cases observed among residents of each market area subtracted from the expected number. A plus means more cases than expected, a minus, less. An asterisk indicates that the difference is statistically significant (p<.01). The expected number is the age-adjusted number of cases that would have occured to area residents if the state rate had applied. The standardized utilization rate gives the age-adjusted rate for each area expressed as a ratio to the state average. Reimbursements above or below expected are estimated by multiplying the average charge for these DRGs for Maryland by the number of cases above or below expected.
[a]Observed minus expected, standardized to state average = 1.00
*Significant (p<.01)

while the efficiency component--(the average reimbursement per cystoscopy not shown in the exhibit)--varies by less than 20 percent. This is typical of most surgical and diagnostic procedures and illustrates the importance of taking the volume into account in the design of cost-containment efforts. The information also raises questions concerning. the effectiveness and efficacy of the various practice styles. Note that in Rumford, nearly 4 percent of enrollees has cystoscopic examinations, while in Waterville and Calais about one percent of enrollees were examined. What are the risks and benefits of these different patterns of use for this technology? We simply don't have a good answer to that question at this time.

Similar tables have been generated from Medicaid and Blue Shield programs for use in feedback to Maine physicians. Under the feedback strategy I suggest, tables such as these should be generated by third-party carriers for all commonly used diagnostic and therapeutic procedures.

Hospital discharge data should also be used to generate age-adjusted utilization experiences for specific causes of admission or surgical procedures. Exhibit 6 illustrates an example of a report useful for feedback

in a DRG-based prospective reimbursement program designed to draw attention to the importance of admission policies. Note that for medical back problems (DRG 243) the rate in the Portland and Lewiston hospital market areas is less than 60 percent of the average, while in the Augusta and Waterville area it is more than 30 percent higher than the average. Portland area residents experienced over 560 fewer cases than expected, based on the state average. The cost implications of the variations in admission rate for DRG-based reimbursement programs are illustrated in the exhibit. Over the two-year period, reimbursements for the Portland population under a DRG reimbursement program would be over $1 million less than expected, based on the state average. In Waterville and Augusta, their combined excess in reimbursements would be $500,000 more than expected. If the Portland use rate were the standard, outlays for medical back admissions in Maine in 1980-81 would have been $7.7 million. If the Waterville rate were the standard, $18.2 million would have been expended. Such displays should be used in DRG programs to bring the variance to the attention of practicing physicians, hospital administrators, and other interested parties. The importance of admission rates in determining expenditures is clearly revealed in this exhibit: more than 63 percent of the causes of hospitalization have admission rates that are more variable than medical back problems.

Dental extractions are among the most variable of causes of admission. Note in Exhibit 6 the more than tenfold range in variation in the standardized utilization rates among the nine individually listed markets in the exhibit. Per capita reimbursements under a DRG program would range from a low of $180 per 1,000 population to a high of $1,860. If the practice style in the Augusta area were the standard for the state, the costs in Maine for this service performed in the in-hospital setting would be about $375,000; if the practice style for Brunswick were the standard, the reimbursements would be ten times higher, or about $3.7 million. Decreasing the use of hospitals for such high-variation procedures offers the potential for large reductions in the cost of hospital care. Reports such as these that identify points where savings can be realized should be used in cost-containment efforts.

A third series of reports is concerned with outcomes. As I have indicated, the practice style factor can play an important role in clinical decision making because the scientific evidence on the consequences of using particular treatments is ambiguous or incomplete. Estimates of survival and complication rates following the use of specific treatments for representative populations are frequently not available, even though they are essential for the evaluation of the common practices of medicine as well as for new technology. Claims data offer an inexpensive means for closing this information gap.

Claims data can be used, for example, for evaluating survival pros-

pects or the probability of a secondary operation following the initial treatment of hypertrophy of the prostate by prostatectomy. I have used the Medicare claims data for such purposes in Maine, finding that the mortality rate in the year following prostatectomy was considerably higher than predicted by most of the published literature. The probability of undergoing a second prostatectomy was also quite high, reaching 13 percent by the end of the fifth year. As illustrated below, such information can help physicians deal with the uncertainties revealed by the practice variation phenomenon, leading to a fuller understanding of the consequences of particular decisions and motivating physicians to take the necessary additional steps to improve the scientific basis of medical practice. Reports based on claims data for analysis of survival and complication rates should become routinely available for technology assessment and the evaluation of the consequences of the natural experiments that derive from the medical practice variation phenomenon.

Is it possible to feedback information to physicians efficiently? Although this idea was first proposed by William Farr and Florence Nightingale well over 100 years ago, recent advances in computer technology, biostatistics, and epidemiology only now make it feasible to produce routinely the reports I am suggesting here. Furthermore, the necessary data are becoming available in many parts of the country. Large, computerized, population-based data files, comprised of hospital discharge records and health insurance claims, now exist in the public and private sectors. Several large states–California, Maryland, Massachusetts, New Jersey, New York, and Iowa–now have state statutes that require hospitals to submit information on the cases they treat to publicly controlled data bases. Public use data bases have been key in our efforts to initiate feedback in the state of Maine. In the late 1970s, primarily under the leadership of David Smith, Alice Russell, and David Soule, the public use of hospital discharge data became a reality with the founding of the Maine Health Information Center.

American corporations, particularly large employers such as the American Telephone and Telegraph company, are beginning to access their own records as a means for managing employee benefit packages. But for purposes of monitoring the activities of local markets, corporate data bases, used by themselves, have severe limitations because, as a rule, no single corporation has enough employees to allow for valid statistical inferences on practice variations in specific hospital markets. Rather, corporations and business or labor coalitions that want to use hospital market data in their cost-containment strategies should. support the development of public data bases on a regional or statewide basis, as exemplified by the Maine Health Information Center. They could also promote information feedback by using their influence as large purchasers of care to insist that third-party carriers publish reports on expenditures and service use rates

in local hospital markets (such as shown in Exhibit 5). Using claims pooled from all Blue Cross accounts, John Putnam of Maine Blue Cross has shown how that organization can provide very important information on variations.

Because of its national coverage and the richness of its data base, the Medicare program offers the best immediate opportunity to implement feedback in all parts of the country. The federal government now requires each hospital to record uniform information on the costs, reasons for hospitalization, and treatments for each hospitalization paid for under the Medicare program. When this information is linked to claims data under the Medicare Part B program and to patient registration files, a registry is created of the medical care events and certain outcomes for virtually the entire population of the United States who are sixty-five years and older. The many problems for public policy concerning the equity and outcome of care that are illustrated by the variation phenomena, as well as the federal government's own need for effective cost containment, lead me to recommend that this very important national resource be used for this purpose.

The Second Step: Dealing With The Effectiveness Problem

My plan for dealing with the effectiveness problem envisions the broad-based feedback of information on practice variations and outcomes targeted to state medical associations, specialty societies, and to individual hospitals and their physician staffs. My hypothesis is that this will result in the reconsideration of the indications for specific services. Some controversies concerning the need for or value of specific practices will be resolved through a critical review of the medical literature or by the application of decision analysis leading to the emergence of more objective standards. (The Rand Project, reported elsewhere in this issue of *Health Affairs*, is an example of a process that may lead to this result.) I expect that such standards, when applied within the context of a review of use rates in specific local markets, will result in a reduction in variation. For other services, reviews of the literature and the use of decision analy sis will identify the controversies and the points of missing data, but cannot, because of the lack of information, lead to a meaningful scientific consensus on the outcome implications. It will, however, greatly refine the debate, identify the critical uncertainties, and should lead the profession to take the necessary steps to obtain more information.

Are these reasonable expectations? Can one expect that state medical societies, specialty organizations, practicing physicians, and academic medicine will pay attention and take action? I am confident that when information is presented in an objective fashion, physicians will respond by accepting responsibility for the outcome implications of the practice

variation phenomenon. I have had occasion to bring specific information on hospital market areas to the attention of the state medical organizations in Vermont, Maine, and Iowa. In Iowa, this may be leading to an important cost-containment program based on provider initiative. In Vermont and Maine, this led to official action by the state medical organizations to endorse the routine feedback of information and to specific proposals to develop programs to deal with efficacy issues raised by the variations.

In the early 1970s when Alan Gittelsohn and I first learned about the large variations in tonsillectomy rates in Vermont, I took the information to the elected officials of the Vermont Medical Society. Without formal program support and principally through the efforts of its past president, Roy Buttles, the society circulated the information on tonsillectomy rates to Vermont hospitals. As a result, Lewis Blowers and Robert Parker, practicing physicians in Morrisville, Vermont, which was identified in the study as the high-rate area, undertook a review of the recent literature and concluded that indication standards for the procedure should be tightened. They convinced their colleagues that hospital policy on the use of tonsillectomy should be changed and that the procedure should be used only after a second opinion was obtained. In subsequent years, the rate for tonsillectomy dropped to less than 10 percent of the rate as first measured. This important example of physician-initiated response to information occurred without economic sanctions and was motivated primarily by concern that local practice patterns should conform to state-of-the-art criteria for recommending tonsillectomy.

When we first learned of the practice variations in Maine, we were invited by Daniel Hanley, who was then executive secretary of the Maine Medical Association and editor of the *Maine Medical Journal*, to write a series of articles setting out the variation phenomenon for Maine physicians. Hanley's initiative, first in publicizing the variations and then in organizing the physicians of Maine into a program to deal with the efficacy issues that variations raise, exemplifies the leadership that practicing physicians can provide. Financial support from The Commonwealth Fund made it possible to undertake a pilot project that initiated the systematic feedback of information and provided the opportunity to demonstrate that practicing physicians are willing to participate in the steps I outlined above to understand the outcome implications of variations. The success and popularity of the pilot project have convinced the Maine Medical Association that it should assume long-range responsibility for running the program of feedback and practice review. Now supported by a grant from the Robert Wood Johnson Foundation, the association is taking steps to secure permanent funds to institutionalize the program.

The strategy is to bring together physicians from market areas with high and low rates for highly variable procedures to discuss the reports

on practice variations. Focusing first on the published literature, can consensus be reached on the significance of the variations and a plan devised for reducing marginally indicated services? If the group concludes that there is underservice within an area, how can the problem be corrected? If consensus cannot be reached, what additional information is required? Can information on survival and complication rates narrow the range of uncertainty? Are prospective studies needed to fill in the gaps? When uncertainties remain, can valid clinical trials be organized to resolve the question of efficacy?

The actions of the physicians convened to study variations in prostatectomy rates answer some of these questions. Although prostatectomy rates varied by a factor of more than 2.5 among hospital markets in Maine, consensus on the appropriate rate could not be reached through review of the literature. A basic uncertainty concerned the survival rate after surgery. Most reports suggest that the mortality rate attributable to this procedure is about 1.2 percent, but this estimate is based on in-hospital experience prior to discharge. Only one report in the literature indicates that mortality rates remain high after discharge, reaching more than 4 percent by the third month following surgery. Moreover, an editorial response to this paper discounted the finding as atypical, explained by patient selection and operative methods. We were able to clarify the situation, using the data most relevant to the physicians of Maine, which was based on their own experience in the treatment of virtually all Medicare patients undergoing prostatectomy in Maine over a two-year period. This study substantiated the longer period of elevated mortality. Indeed, the data show that for one subgroup, the risk of death within a year was about 40 percent; in reviewing this information, a consensus emerged among the physicians that these patients were better treated by more conservative methods.

The physicians have also studied the evidence underlying their assumptions about the benefits of prostatectomy, particularly the expected gains in the quality of life. Again finding gaps in the literature, they have been motivated to undertake a prospective study to ascertain the objective as well as subjective responses of their patients to the surgery. Currently being designed with the assistance of academically based collaborators, this study will represent the first large-scale, population-based follow-up to document nonfatal outcomes associated with the use of this procedure. When completed, the study will help all physicians make better decisions about when to recommend prostatectomy as well as pinpoint remaining uncertainties that may need to be settled by a clinical trial.

The active involvement of Maine physicians in examining practice variations is indicative of the response to be expected from most medical practitioners. The special status of the medical professional derives partly

from the exercise of collective responsibility for understanding illnesses and the consequences of alternative therapies, and for helping patients realize the medical care they truly want. The professional uncertainties and disputes about outcomes that underlie some examples of variations present a direct intellectual challenge to practicing physicians as well as academic-based researchers. They also indicate that past efforts to distinguish the scientific from the unscientific claims concerning effectiveness have not been sufficient; greater efforts are needed to base clinical choices more solidly on sound estimates of outcome probabilities and on values that correspond closely to patient preferences. My experiences indicate that information on practice variations, when used in a program of feedback that includes epidemiologic, biostatistical, educational, and financial support, motivates practicing physicians to take the necessary steps to improve clinical decision making.

The uncertainties about clinical outcomes are particularly important for academic medicine because of its special responsibility for the science of medicine. If, as I propose, the feedback of information on service use variation is broadened to include the populations served primarily by prominent teaching institutions, interest in the significance of the variation phenomenon may be considerably enhanced. When this is accomplished for the Boston and the New Haven markets, some very interesting practice style variations will emerge as contributors to the more than twofold difference in per capita expenditures exhibited in Exhibit 4. Intellectual curiosity and the need to justify such differences in costs should lead naturally to sophisticated efforts to explicate the significance of the differences in practice styles.

There are other reasons why the variations should be of interest to academic medicine, not the least being their responsibility for the training of new physicians. In my opinion, the state of intellectual confusion on the rational use of medical services evidenced by the monitoring of local market performance calls upon academic medicine to increase the attention and support it gives to the disciplines involved in improving clinical decision making–to clinical epidemiology, biostatistics, and clinical decision analysis. An important topic for the research agenda is to improve methods for evaluating health care outcomes, particularly means for measuring functional status. Today's dilemmas stem, in part, from advances in biomedical research, and the natural next step is to improve the quality of research into the impact of these investments.

There are also implications for medical education. Medical students need more extensive and better education in the methods of evaluating clinical decisions and their outcomes so they may assess for themselves the strengths and weaknesses of the various practice styles they will encounter in the course of their clinical training and prepare for their own contributions to resolving clinical uncertainties as practicing phy-

sicians.

Practicing physicians and their medical associations cannot act without broad-based support. Private philanthropy is playing a crucial role by providing leadership in mobilizing opinion on the importance of the practice variation problems. Examples include The Milbank Memorial Fund's investment in the development of clinical epidemiology and the support of The Commonwealth Fund, The Robert Wood Johnson Foundation, and The John A. Hartford Foundation to find solutions in basic research as well as demonstration projects such as the Maine Medical Association project. With few other exceptions, however, there is currently little support in the private or public sector for the research needed to establish the outcome value of common medical practices.

The federal government's lack of effective policy is noteworthy. The National Center for Health Care Technology, whose agenda was technology assessment, has been abandoned and funding for the National Center for Health Services Research, which has supported much of the research upon which my proposal is based, is minuscule in comparison to the need. The Health Care Financing Administration invests little or none of its research resources in projects concerned with the health outcome value of the services it pays for.

The failure of technology assessment to attract public support is all the more surprising in view of the implications of the uncertainty concerning surgical mortality. For example, if the conservative practice style for prostatectomy observed in some New England areas were the national norm, the number of postoperative deaths in the United States would be about 1,900; under the liberal style, the number would be about 6,800, suggesting that under the high-rate strategy about one percent of American males over age sixty-five would die postoperatively. Most prostatectomies are paid for by the federal government. The public interest is served by a better understanding of the implications of the variations. The responsibility for furthering research into the outcome implications seems to rest in part with the federal government because many of its activities promote the public's use of health care.

The Third Stepr Dealing With The Cost-Containment Problem

Many hospitalized patients can be effectively and safely treated in the ambulatory setting; the problem is knowing who they are. The shift of such patients to the ambulatory setting will neither disrupt the patient-physician relationship nor have a significant negative economic impact on physicians. Given the current imperatives to contain the costs of medical care and reallocate resources to more productive ends, it should be in most peoples' interest to reduce the use of hospitals for marginally indicated causes of admission and to translate the reduction in demand