

Dartmouth College

Dartmouth Digital Commons

Dartmouth Scholarship

Faculty Work

2023

The Wennberg Anthology

John E. Wennberg
Dartmouth College

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>



Part of the [Epidemiology Commons](#)

Dartmouth Digital Commons Citation

Wennberg, John E., "The Wennberg Anthology" (2023). *Dartmouth Scholarship*. 4311.
<https://digitalcommons.dartmouth.edu/facoa/4311>

This Other is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

The Wennberg Anthology

John E. “Jack” Wennberg, MD, MPH

Founder and Emeritus Director of The Dartmouth
Institute for Health Policy & Clinical Practice

Emeritus Professor of The Dartmouth Institute and
of Community and Family Medicine



The Wennberg Anthology

John E. “Jack” Wennberg, MD, MPH

Founder and Emeritus Director of The Dartmouth Institute for Health Policy & Clinical Practice (as the Center for the Evaluative Clinical Sciences at Dartmouth)

Emeritus Professor of The Dartmouth Institute and of Community and Family Medicine

Table of Contents

An Introduction	3
The Early Papers	4
Small Area Variations in Health Care Delivery	5
Health Care Delivery in Maine I: Patterns of Use of Common Surgical Procedures.....	7
Health Care Delivery in Maine II: Conditions Explaining Hospital Admission.....	8
Health Care Delivery in Maine III: Evaluating the Level of Hospital Performance	9
A Test of Consumer Contribution to Small Area Variations in Health Care Delivery	10
Changes in Tonsillectomy Rates Associated with Feedback and Review	11
A Report from the Cooperative Health Information Center of Vermont	12
A Small Area Approach to the Analysis of Health System Performance	13
Professional Uncertainty and National Priorities for Use of Resources	15
PSRO and the Relationship among Health Need, Elective Surgery and Health Status	16
Small Area Variations in Health Care Delivery: A Critique and Response	17
The Prostate Papers: Outcomes Research and the Pathway to Shared Decision-Making.....	18
The Preventive Versus the Quality of Life Theory of Surgery for BPH.....	19
What are the Gains in Quality of Life?	20
Shared Decision-Making and the BPH Interactive Video Disc	21
Learning What Patients Want and Which Rate is Right.....	22
The Open Versus TURP Controversy	23
Collaboration with the American Urologic Association	25
The BPH Treatment Outcomes Study	26
BPH and the PORT Strategy for Patient-Centered Outcomes Research.....	27
The Prostate Port and Early Stage Prostate Cancer	28
PORTS and Politics: The Rise and Fall of Medical Effectiveness Research	31
The Foundation for Informed Medical Decision-Making.....	32



The Epidemiology of Medical Care	33
How Much Variation is There?.....	35
The Boston/New Haven Trilogy	37
The Dartmouth Atlas of Health Care.....	40
Is More Better?	41
Race, Poverty, and Utilization	45
Evaluating Provider Efficiency in Managing Chronic Illness.....	46
Recent Research: Risk Adjustment and the Observational Intensity Bias.....	48
The Policy Papers	51
The Need for Assessing the Outcomes of Common Medical Practices	52
Professional Uncertainty and the Problem of Supplier-Induced Demand	53
Should the Cost of Insurance Reflect the Cost of Use in Local Hospital Markets?.....	54
Dealing with Medical Practice Variations: A Proposal for Action	55
Small Area Analysis and the Medical Care Outcome Problem	56
What is Outcomes Research?	57
Are Randomized Controlled Trials Controlled? Patient Preferences and Unblind Trials	58
Innovation and the Policies of Limits in a Changing Health Care Economy	59
Avoiding the Unintended Consequences of Growth in Medical Care: How Might More Be Worse?.....	60
The Marketplace in Health Care Reform: The Demographic Limitations of Managed Competition	61
Finding Equilibrium in the U.S. Physician Supply	62
Benchmarking the U.S. Physician Workforce.....	63
Prioritizing Oregon’s Hospital Resources.....	64
Regional Professional Foundations	65
The Randolph Plan	66
The Hanover Experiment	68
Geography and the Debate Over Medicare Reform	69
Toward the Tipping Point: Decision Aids and Informed Patient Choice	70
Extending the P4P Agenda	71
An Agenda for Change: Improving Quality and Curbing Health Care Spending. Opportunities for Congress and the Obama Administration	72
Editorials, Commentaries, & Perspectives.....	73



An Introduction

In 1967, after finishing my postgraduate training at Johns Hopkins in internal medicine and epidemiology, I took a job at the University of Vermont in Burlington as Director of the Northern New England Regional Medical Program (RMP). The Vermont program was one of some fifty RMPs that blanketed the country as part of President Johnson's Great Society program. The idea behind the RMP was that advances in biomedicine had so improved the outcomes of major killer diseases that it was critical to ensure that all Americans, not just those fortunate enough to live near an academic medical center, had access to these services.

Epidemiologists are interested in what happens to groups of people: for example, how many have heart attacks (incidence) and what happens to patients (outcomes) according to the treatment they receive (prognosis). It was thus quite natural for me to want to use the tools of epidemiology to provide population-based information about the distribution of health care resources and the utilization of services among Vermont communities; after all, good planning for improvement requires knowledge about the current status of the health care system.

Alan Gittelsohn, a biostatistician from Johns Hopkins who had also been my teacher, and I developed a strategy we called the small area analysis of health care delivery. The method defined the geographic boundaries of local health care markets, based on where patients actually went for their care, and described the per capita use of resources and services for the resident populations.

Our first Vermont small area analysis (which we published in a [1973 article in Science](#)) brought a big surprise. While we had expected to find a rural health care system characterized by underservice, we found instead a typology of care characterized by vast variations in the deployment of resources and the utilization of services among neighboring communities, without apparent rhyme or reason. The results forced me to re-examine the policy assumptions behind the RMP and a good deal else about the U.S. health care economy. It was evident that the problems facing regional and local health care markets were much more profound than the barriers to diffusion of new technology the RMP was designed to overcome.

I have spent most of my career studying the variation phenomenon. In a recent book, [Tracking Medicine: A Researcher's Quest to Understand Health Care](#), I chronicled this research and the evolution of thinking on the causes and remedies for unwarranted variation. Several of my colleagues and students have asked if I would make our research papers—some of which are hard to find and some not yet published—available to a wider audience. This website is my effort to do this. I provide introductory remarks for each paper in an attempt to clarify the context, relationship to previous work, and the role the ideas in these papers played in building an understanding of practice variation. Over the years I have had the good fortune of working together with colleagues who have shared my fascination with the story of practice variation, and who have contributed in fundamental ways to the ideas and concepts presented in these papers.



The Early Papers

Published between 1973 and 1980, these papers describe research undertaken in Vermont and Maine that first developed the methods of small area analysis; uncovered the extent of the variation phenomenon; tested theories as to its origins; and explored the implications of the professional uncertainty hypothesis and supplier-induced demand.

1. [Small Area Variations in Health Care Delivery](#)
(*Science*, 1973)
2. [Health Care Delivery in Maine I: Patterns of Use of Common Surgical Procedures](#)
(*Journal of the Maine Medical Association*, 1975)
3. [Health Care Delivery in Maine II: Conditions Explaining Hospital Admission](#)
(*Journal of the Maine Medical Association*, 1975)
4. [Health Care Delivery in Maine III: Evaluating the Level of Hospital Performance](#)
(*Journal of the Maine Medical Association*, 1975)
5. [A Test of Consumer Contribution to Small Area Variations in Health Care Delivery](#)
(*Journal of the Maine Medical Association*, 1977)
6. [Changes in Tonsillectomy Rates Associated with Feedback and Review](#)
(*Pediatrics*, 1977)
7. [A Report from the Cooperative Health Information Center of Vermont](#)
(CHIC, 1974)
8. [A Small Area Approach to the Analysis of Health System Performance](#)
(DHHS, 1980)
9. [Professional Uncertainty and National Priorities for Use of Resources](#)
(DHEW, 1977)
10. [PSRO and the Relationship among Health Need, Elective Surgery and Health Status](#)
(Conference Proceedings, Boston University Medical Center, 1973)
11. [Small Area Variations in Health Care Delivery: A Critique by Francis D. Moore, MD](#)
(*Journal of the Maine Medical Association*, 1977)
12. [The Authors Respond, Alan Gittelsohn, PhD and John E. Wennberg, MD](#)
(*Journal of the Maine Medical Association*, 1977)



Small Area Variations in Health Care Delivery

- Wennberg JE, Gittelsohn A. [Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making](#). *Science*. December 14, 1973 1973;182(4117):1102-08.

This was our first published report on variations in health care delivery among Vermont hospital service areas.

Our “small areas”—13 geographically distinct Vermont hospital service areas—were defined empirically, based on an analysis of place of residence and the hospital most commonly used. We sought to assemble information on each sector of care: inpatient, ambulatory, and nursing homes. Measures were developed to compare health care delivery in the 13 areas on a per capita basis including resource inputs (expenditures, physician labor by specialty, hospital and nursing home beds, and personnel); utilization of services (hospitalizations by cause, surgical procedures, diagnostic tests, nursing home admissions); and need and outcomes, as measured by population age structure and mortality.

The key to understanding our small area method is that it is population-based: resource inputs, utilization and outcomes are for the residents of the hospital service area, no matter where the services are obtained. However, since the vast majority of care in each area is provided by local physicians, hospitals, and nursing homes, the variation we uncovered among hospital service areas was deeply influenced by local practice patterns and local medical supply.

Here is a brief synopsis of what we found:

- There were extensive variations among Vermont hospital service areas in virtually all aspects of health care delivery. Expenditures for hospitals varied twofold; for nursing homes, fivefold; and Medicare reimbursements for physicians, threefold. The supply of physicians, hospital beds and nursing home beds exhibited similar variation. Hospitalization rates for most causes of admission varied two to threefold; common surgical procedures varied tenfold among areas for tonsillectomy; fourfold for hemorrhoid surgery and threefold for prostatectomy, hysterectomy and gall bladder surgery.
- Physician supply characteristics were correlated with variation in demand (utilization). The specialty mix of physicians correlated with the mix of services provided to a defined population: populations living in areas with more surgeons per 10,000 had more surgery per 10,000 at all levels of complexity; areas with more general practitioners doing surgery had higher rates of less complicated surgery; and populations living in hospital service areas with more internists per 10,000 underwent more diagnostic tests, but less surgery.
- The variation in physician supply (on a per capita basis) bore no apparent relationship to population need. Hospital service areas with older populations and lower per capita income had fewer physicians per capita. However, greater spending for hospitals and physician services showed no significant correlation with age-adjusted mortality and perinatal mortality.



In an evaluation of the variation, we suggested the variation was linked to professional uncertainty about the best way to practice medicine:

“The variations are more likely to be associated with differences in beliefs among physicians concerning the indications for, and the efficacy of, the procedure (than differences in the incidence of disease).”

And uncertainty about best practice was linked to poor clinical science:

“An important reason for uncertainty is that few prospective clinical trials have been performed. Because the outcome of one type of service compared to another (or to none at all) is often unknown, the variations in therapeutic or diagnostic procedures observed among different Vermont communities cannot be strictly evaluated.”

As a consequence, we concluded, it wasn't clear that more is better:

“(G)iven the magnitude of these variations, the possibility of too much medical care and the attendant likelihood of iatrogenic disease is presumably as strong as the possibility of not enough service and unintended morbidity and mortality.”

Over the years, pursuit of answers to the “which rate is right” question became a central theme in the effort to understand practice variation.

Our small area analysis also illustrated a connection between medical variation and unintended consequences of public regulation, such as:

unwarranted transfer payments from regions with low spending to regions with high spending:

“Variations in expenditures, sustained in large part through third party payment mechanisms, pose questions of equity, since the price of insurance is not adjusted to reflect these differences.”

poorly informed public regulatory decisions affecting per capita expenditures and bed supply—errors that resulted in increasing rather than reducing variation. These errors, we suggested, could be avoided if population-based data had been available (and used) by the Hill-Burton Program (in determining need for beds) and the Price Commission (in regulating prices).

We concluded that “(p)opulation-based health information systems ... are an important step in the development of rational public policy.”

Over the years, my colleagues and I would continue to emphasize the value of the medical care epidemiologic perspective in achieving accountability for performance in the private as well as the public sector. We would extend small area studies to other New England states, to Iowa, California and, eventually, to the entire United States as part of the [Dartmouth Atlas project](#).



Health Care Delivery in Maine I: Patterns of Use of Common Surgical Procedures

- Wennberg JE, Gittelsohn A. [Health care delivery in Maine I: patterns of use of common surgical procedures](#). *J Maine Med Assoc*. May 1975;66(5):123-30, 149.

In the early 1970s, John Putnam and David Soule visited Vermont to tell us about their own data project in Maine and asked us to apply our small area analysis method to Maine. Since we were anxious to find an opportunity to see if the Vermont findings could be replicated elsewhere, we gladly agreed. The results were reported in a series of three articles published in the *Journal of the Maine Medical Association* in 1975.

This article demonstrated a pattern of surgical variation in Maine similar to that seen in Vermont. It developed further the argument that variation itself helps us understand the important role physician opinion plays in influencing utilization. The case in point was the idiosyncratic pattern of use of five common surgical procedures in the five largest regions of Maine—the earliest example of what we came to call the “surgical signature” of a community. [Figure 4, Maine I]

We argued that the surgical signature phenomenon results from differences in local medical opinion that in some cases can only be resolved by empirical studies of outcomes.

In an accompanying editorial, Dr. Daniel Hanley, the editor of the *Journal of the Maine Medical Association*, urged his fellow physicians to respond to the challenge of practice variations:

“All this will require a series of detailed looks by those who know the situation best—the physicians themselves—this means time and effort, but the rewards are great—a better understanding of the decision-making process ... The publication of this data will raise some hackles, but it will also give thoughtful doctors cause to look at a new dimension and to begin to talk about how to measure objectively the improvement in medical care.”

With these remarks, Dan set in motion a collaboration between our research group and practicing physicians in Maine that over some twenty or so years would clarify the clinical reasons for variation and, through outcomes research, improve the scientific and ethical basis for surgical decision-making.



Health Care Delivery in Maine II: Conditions Explaining Hospital Admission

- Wennberg JE, Gittelsohn A, Soule D. [Health care delivery in Maine II: conditions explaining hospital admission](#). *J Maine Med Assoc*. Oct 1975;66(10):255-61, 269.

In this study, we asked how the pattern of variation differed according to the condition for which the patient was admitted to hospital. For a few conditions such as pregnancy and congenital anomalies, we found that the hospitalization rates varied little among Maine hospital service areas. Why? It seemed that these “low variation” conditions were easily and unambiguously diagnosed and that almost all physicians agreed that the patient should be hospitalized. Because of this professional consensus, we argued that the hospitalization rates were largely determined by the birth rate and the incidence of congenital anomalies. However, for most causes of admission, the hospitalization rate did not follow the low variation pattern. Most hospitalizations were for “high variation” conditions and the variation seemed unwarranted—it wasn’t easily explained on the basis of illness.

The findings led us ask, how much variation is there? What proportion of the hospital case mix corresponds to a model of hospital demand in which the incidence of disease is the prominent predictor of hospitalization? What proportion seems susceptible to supplier influence? In Maine, the vast majority—over 80%—were high variation; only 18% of hospitalizations fit the low variation model. Subsequent research, using more sophisticated measures of variation and more comprehensive definition of case mix, provided approximately the same answer.



Health Care Delivery in Maine III: Evaluating the Level of Hospital Performance

- Wennberg JE, Gittelsohn A, Shapiro N. [Health care delivery in Maine III: evaluating the level of hospital performance](#). *J Maine Med Assoc*. Nov 1975;66(11):298-306.

The efficiency of hospital care has traditionally been measured by hospital-based indicators such as average length of stay in hospital, average cost per case or patient throughput per bed, all under the assumption that these measures reflect the underlying use of resources and services by the population the hospital serves. In the [article in Science](#), we showed that in Vermont such measures were unreliable indicators of underlying differences in per capita resources inputs or utilization.

In this paper, we showed a similar pattern for Maine: for most causes of hospitalization and for most surgical procedures, it was the volume of care (i.e. the population-based admission rate), not the average use of resources per admitted patient (i.e. cost per case, length of stay), that determined most of the variation in spending or use of hospital beds per capita among local medical markets. Similarly, occupancy rate, long used as both an indicator of need for more beds (when it is high) or inefficiency (when it is low), was shown to be uncorrelated with per capita bed supply, the “bottom line” policy variable. Public regulatory decisions, undertaken without benefit of population-based indicators, increased the inequality in distribution of resources among neighboring areas, and established income transfers through insurance mechanisms. The article concluded, as did the *Science* article, that *a direct, epidemiological approach to evaluation of performance is necessary if basic issues concerning medical necessity and distributional equity are to be identified*.

The paper also presented case histories of how population-based data could be used in distinguishing inappropriate use of hospitals from shortage due to low bed supply in evaluating the need for additional beds. One of my favorites is The Case Study of Need for Coronary Intensive Care Beds, which grew out of the Vermont RMP’s regional disease management program. It is an example of how direct measures of population use of resources were used in deciding whether more resources were needed. The use of such population-based “best practices” benchmarks to evaluate resource needs was to become a central concern of our group in the 1990s.



A Test of Consumer Contribution to Small Area Variations in Health Care Delivery

- Wennberg JE, Fowler FJ, Jr. [A test of consumer contribution to small area variations in health care delivery](#). *J Maine Med Assoc.* Aug 1977;68(8):275-9.

The small area studies in Vermont and Maine challenged the conventional wisdom that the use of health care was primarily determined by illness and the socioeconomic circumstances of patients. In 1973, we were able to formally test this hypothesis through a household interview study that sampled patients living in six Vermont hospital service areas that varied substantially in their hospitalization rates, their surgery rates, and expenditures. The survey obtained information on factors known to predict the behavior of individual patients in seeking care. The primary predictor is illness. Others include income and educational levels, health insurance coverage, cultural background, and availability of services. We asked two critical questions: Were there differences in these factors among the six hospital service areas? And, if so, did they correlate with the amount of care consumed?

In contrast to the twofold or greater difference in use of medical care, the populations of the six Vermont medical communities were remarkably homogeneous in factors that predict the individual's need for and ability to consume care. For the few examples where minor differences were found among areas, they weren't correlated with utilization and expenditures.

The similarity among the populations in illness rates and socioeconomic factors associated with use of care predicted that the residents of the six areas would not differ in their propensity to seek care. Their rate of contact with their physicians confirmed this expectation: On an annual basis, between 71% and 77% of the population in six hospital service areas contacted their physician at least once. The differences in hospitalizations, surgery and expenditure seemed better explained by *post-access provider behavior* that varied according to characteristics of the local health care system.



Changes in Tonsillectomy Rates Associated with Feedback and Review

- Wennberg JE, Blowers L, Parker R, Gittelsohn AM. [Changes in tonsillectomy rates associated with feedback and review](#). *Pediatrics*. Jun 1977;59(6):821-6.

When the information on the variation in surgery rates for 1969 first became available, I shared the information with Dr. Roy Buttles, the president of the Vermont Medical Society, who circulated it among the members of the Society's utilization review committee. By 1973, the tonsillectomy rates had declined dramatically in most Vermont hospital service areas and the variation in rates among areas measured by the coefficient of variation had decreased from .67 to .40.

The rates in the Morrisville hospital service area showed the greatest decline, dropping nearly 90% by 1973. Upon learning of the high rate, Dr. Lewis Blowers, a general surgeon, and Robert Parker, a pediatrician, took it upon themselves to initiate an active review of all candidates for tonsillectomy. When I learned about the drop in rates, I visited with Dr. Blowers and Parker and we agreed to write a paper that would trace the rates of tonsillectomy among the 13 Vermont hospital service areas over a five-year period (1969-1973) and discuss what had happened in Morrisville.

This article provides important "first-hand" testimony concerning the influence of physicians on utilization. Our co-authors Blowers and Parker attributed the rapid decline in rates to a change in their own (and their colleagues) practice styles in response to feedback and their subsequent review of the scientific literature. Their story convinced many in the medical community as well as some social scientists of the importance of supplier-induced demand—and indicated that feedback on variation could lead to rapid change in practice patterns. It also showed that physicians were willing to take responsibility for addressing practice variation. Experience in Maine would further strengthen my belief in the importance of medical leadership.



A Report from the Cooperative Health Information Center of Vermont

- Gittelsohn AM, Robbins J, Gillim R, Senning J, MacPherson B, Hickcox P, Provost K, Herr D, Leveton J, Higgins B, Zablocki R. Vermont Surgery Study, 1969-1971: [On the Incidence of Tonsillectomy and Other Common Types of Surgery](#). Cooperative Health Information Center of Vermont, 1974.

The Regional Medical Program proved an unstable home base for the population-based health information system that was at the core of the Vermont strategy. In the early 1970s, the Cooperative Health Information Center of Vermont was created as an independent, not-for-profit organization whose sole responsibility was to maintain and further develop the population-based health information system begun by the RMP to provide periodic feedback on performance to physicians, hospitals, nursing homes, regulators and the general public. The president was Deane C. Davis, a former governor of Vermont, and the board of directors included prominent leaders from both political parties as well as David Bazelon, chief justice of the United States Court of Appeals for the District of Columbia Circuit in Washington, DC, who was a leading advocate for informed consent.

Our effort to institutionalize the Vermont health information system and to design feedback reports was based on the assumption that regulators would respond (as we said in *Science*) by developing “explicit strategies to deal directly with inequalities and uncertainty concerning the effectiveness of health care delivery”. Unfortunately, a shared understanding of the causes and remedies for unwarranted variation, a key requirement for rational public policy for health, never emerged in Vermont, neither among regulators nor, despite the efforts of the Vermont Medical Society, among providers, in particular the leaders of the academic medical center in Burlington. Eventually, the energy to maintain the database dissipated from lack of effective use and in 1983, the Cooperative Health Information Center of Vermont closed its doors.

Alan Gittelsohn served as the Center’s first director. Together with Jennifer Robbins, Roger Gillim, John Senning, Brian MacPherson, Patricia Hickcox, Karen Provost, David Herr, Jeff Leveton, Barbara Higgins, and Rita Zablocki, he wrote a report entitled “Vermont Surgery Study, 1969-1971: On the Incidence of Tonsillectomy and Other Common Types of Surgery”. It stands as an important example of the use of medical care epidemiology to provide feedback on surgical performance.



A Small Area Approach to the Analysis of Health System Performance

- Wennberg JE and Gittelsohn AM. [A Small Area Approach to the Analysis of Health Systems Performance](#). (Monograph). U.S. Department of Health and Human Services, Public Health Service, Health Resources Administration, HRP-0102101, 1980; DHHS publication no. (HRA)80-14012.

Throughout the 1970s, Alan Gittelsohn and I continued to develop the methods of small area analysis and apply them to measure the performance of the health care systems in New England. By the end of the decade, we had extended the patient origin analysis across all six New England states, resulting in the division of this territory into 193 geographically distinct hospital service areas. Thanks to generous support from the federal government—and the access this support provided to the Medicare data—we were able to develop new methods for using Medicare claims database to describe the health delivery system; we then applied these methods and data in a series of studies to look at variation in health care delivery throughout New England.

Alan and I were encouraged by the several agencies that funded our work to write a comprehensive review of our methods research and its principle findings and to consider their policy implications. The result was a three-part book that the federal government published as part of its Health Planning and Technology Series. Difficult to find and never marketed, the book remains the best description of the methodologies we developed the 1970s. It provides early examples of the use of Medicare claims data to benchmark performance and shows how population-based data could be used in making decisions regarding the hiring of physicians, the building of beds and in setting prospective budgets in ways that reduce transfer payments among regions.

Part One is devoted to methodology.

Chapter II describes how a small-area analysis is undertaken; how the demography of hospitalizations within a region can be described so that larger geographic regions can be subdivided into local hospital market areas; how indicators of performance describing per capita resource allocations are developed; and how utilization of service is measured. The chapter also deals with several statistical issues that are important in the interpretation of the data.

Chapter III deals with problems of the quality of data available in the hospital discharge abstract. Studies on the consistency of recording practices are reviewed to give the reader an understanding of the limits on the interpretation of hospital utilization data. We concluded that while demographic data, such as age, sex, place of residence and surgical procedure codes are reasonably accurate, the coding of the cause of admission is extremely variable and could not be relied upon at a level of detail beyond major disease classification. Although coding practice has improved with time, significant difficulty with interpretation of diagnoses recorded in large databases remain.



Chapter IV describes work that served as the basis for later research in medical outcomes and our public reports through the Dartmouth Atlas project. We extended the basic methods of small area analysis to the data maintained by the national Medicare program. The data files we used for this purpose are the MedPAR (Medicare Provider Analysis and Review) files, a 20% sample of all inpatient stay records paid for under the Medicare program. The MedPAR files contain utilization information from bills, demographic information about enrollees, and provider information. Data from the Medicare Part B Program, which have the virtue of being potentially available throughout the United States, are also used in Chapter IV. They also have some important advantages over the Maine, Rhode Island, and Vermont hospital data in that information on charges and reimbursements is available. More importantly, the Medicare data include date of death after discharge from hospital, so that survival can be measured. Thus, this source of data can be used to connect medical care events to their outcomes.

Part Two of the book is a description of the epidemiology of medical services in New England.

Chapter V shows how the data have been used to systematically measure the distribution of hospital resources across the six New England states. It also describes the cross-sectional pattern of variation in rates of use of services to show that some surgical procedures or hospitalizations have relatively little variation while others have a great deal.

Chapter VI investigates the consistency of the differences in rates of use of services over time, as well as the long-range trends in expenditures for particular surgical procedures, such as tonsillectomies or cesarean sections.

A picture of the health care system emerges in which the constituent hospitals have strikingly different performance records with regard to impact on population rates of use of medical technology and expenditures.

In Part Three, we give examples of a number of different applications of the data.

Chapter VII explores how epidemiological information can be used for the planning, management or regulation of the hospital system. Theoretical and actual applications of data in facility planning, in Certificate of Need decisions, and in the development of state health plans are discussed and examples given. Applications in insurance or price regulation are also presented, including the use of data to develop “parity” prospective budgeting: based on either Medicare data or the hospital discharge data, budget estimates can be developed for each hospital in a region, such that if these budgets were realized, the variations in hospital expenditures or Medicare reimbursements per capita would be substantially reduced. The importance of using these methods to achieve a more equitable distribution of resources is highlighted by the use of the data to describe the cross-subsidizations between communities that result from insurance or taxation policies that do not take historical patterns of per capita expenditures into account. A final section discusses how hospital administration may use the data for planning of physician staffing and for cost containment through review of employment practices. Our case study of the successful use of benchmarking of physician labor input to make decisions on the need for physicians is presented.



Professional Uncertainty and National Priorities for Use of Resources

- Wennberg JE. [Professional Uncertainty and National Priorities for Use of Resources](#). In: DHEW, editor, *Priorities for the Use of Resources in Medicine*. Washington, D.C.: U.S. Government Printing Office; 1977. p. 301-318.

In 1976, I was asked by the organizers of an NIH conference on health care planning to address whether national priorities were needed for health resource planning and, if so, how should they be developed. The conference was motivated by the recent passage of the Health Planning and Resource Development Act which established a national program charged by the U.S. Congress to achieve “equal access to quality care at reasonable cost” and to improve the health status of the population. While the implementation of Medicaid went a long way to assure access to care for low-income families, the studies in Vermont and Maine had made it clear that after access, populations with similar needs received very different amounts and types of care, depending on where they lived and the doctors and hospitals they used. I was quite pessimistic that health planning, decentralized as it was into some 200 regional planning agencies, could assure that the health care in its region of responsibility was “high quality” and the services provided had a reasonable prospect of improving health.

In my talk, I focused on two drivers of undisciplined growth in American medicine. The first is the accelerating rate of introduction of new technologies that are of uncertain value because they have not been adequately evaluated in terms of impact on health. The second is the accelerating growth in specialists and decrease in primary care, a trend that is producing increased numbers of physicians who advocate use of greater amounts of technology and who, by virtue of their narrow focus, are unable to satisfy the needs of patients for access to a physician when they become sick.

As remedy, I suggested the need for a national program to improve the scientific basis of clinical decision-making by evaluating the outcomes of every day medical practice; while this suggestion had no immediate impact, a decade later we had the opportunity to develop models for how such research can be conducted, as our group undertook an evaluation of the treatment of prostate disease in Maine.

I also suggested that federal subsidies for educating the physician workforce be redesigned to favor the expansion of primary care, again without impact. However, questions about the appropriate size and specialty composition of the physician workforce became a recurrent theme in subsequent policy research by the Dartmouth team.



PSRO and the Relationship among Health Need, Elective Surgery and Health Status

- Wennberg JE. [PSRO and the Relationship among Health Need, Elective Surgery and Health Status](#). Conference Proceedings, *Perspectives on Health Policy*, Boston University Medical Center, 1973.

In October of 1972, Congress passed legislation designed to rein in rapidly increasing medical costs by focusing physician peer review on costly overuse of the acute care hospital. In each state, Professional Standards Review Organizations, or PSROs, were established to review the elective admissions to hospital by developing clinical standards for use in a “pre-admission” certification process to weed out unnecessary care. But it soon became apparent that there was little consensus on the clinical standards that would define medical necessity, and massive resistance soon developed on the part of the profession for the implementation of the certification process.

I was invited to speak at a conference on pre-certification at Boston University, and I took the opportunity to interpret the significance of the Vermont findings on practice variation for the PSRO program, in particular the implication for the development of consensus based standards of care. The studies—as reviewed in this article—had convinced me that for most causes of hospitalization there was no underlying professional consensus on “best practices” that could support meaningful clinical standards on which to base the decision to hospitalize.

As a first principle, I suggested that medical necessity be interpreted in terms of the relationship between health care use and health outcomes, and that the PSRO program should become an active agent for reducing scientific uncertainty about this relationship. Instead of focusing on consensus based guidelines of dubious value, the PSRO program should organize the profession to respond to geographic variation in practice patterns, undertake the studies required to understand their outcome implications, and implement clinical solutions to reduce variation based on this understanding.

While these ideas had little impact on the PSRO program (or its successor organization, the “professional review organization”), they bore fruit in Maine.



Small Area Variations in Health Care Delivery: A Critique and Response

- Moore FD. [Small area variations in health care delivery. A critique](#). *J Maine Med Assoc*. Feb 1977;68(2):49-52.
- Gittelsohn A, Wennberg JE. [The authors respond](#). *J Maine Med Assoc*. Feb 1977;68(2):53-57.

Our practice variation studies challenged the conventional wisdom about how health care markets work and our methods and findings came under close scrutiny and sometimes heated criticism. In my book, *Tracking Medicine*, I described some of the difficulties we had in publishing our earlier articles in “main line” medical journals. The critique by Francis Moore illustrates the intensity of emotion that our findings could evoke among those in high places. At the time of this exchange, Dr. Moore was the Mosely Professor of Surgery at Harvard Medical School and Chief of Surgery at the Peter Bent Brigham Hospital and the principle investigator of a national surgical study—The Study on Surgical Services for the United States, or SOSSUS. I had met Dr. Moore while I first came to Harvard and we talked about the Vermont and Maine findings. A bit later, Dr. Dan Hanley, the editor of the *Journal of the Maine Medical Association*, let me know that Dr. Moore had submitted a critique, and invited Alan and me to respond.

I have included our point-by-point response and a final note from Moore dated January 12, 1977. Judging from his final note, our most effective response was to his concern that we had failed to include obstetrical delivery and reduction of fractures in our description of variation, conditions in which, to quote Moore “the capricious whim of hospitals, surgeons or physicians has absolutely nothing to do with local incidence”.

In fact we had already published (in [Maine II](#)) some of the data Moore wanted. In response, we took the opportunity to show graphically (Figure 1) that delivery-related conditions showed little variation as Moore had hypothesized. The graph seemed to satisfy him because, as he said in his final note, it “clearly indicates that ‘provider behavior’ is not the only source of ‘small area variation in health care delivery’” (which, of course, we had never said).

But hospitalizations for fractures did not meet his expectation for low variation. The reason? Many types of fractures are treated outside of the hospital and the percentage of those so treated varies from place to place. But some life-threatening fractures, notably hip fractures, do indeed pass the low variation test. In subsequent work, we used the pattern of variation in hip fracture as a benchmark for low variation conditions.

Subsequent research would extend the study of variation in hospitalization rates to the entire case-mix to show that most causes of admission follow the high variation pattern.



The Prostate Papers: Outcomes Research and the Pathway to Shared Decision-Making

Following the publication in the *Journal of the Maine Medical Association* of our article describing surgical variation, Dan Hanley, the editor of the journal, organized the Maine Medical Assessment Program to bring physicians together to discuss why medical practice differed so much from one community to another. The collaboration with urologists focused on treatment patterns for benign prostatic hyperplasia, or BPH. In some medical communities in Maine, more than 50% of men had surgery for this condition, while in others, less than 15% did.

In discussing the variation, it soon became apparent that the urologists disagreed among themselves on the fundamental reason for recommending surgery. Most believed in the preventive theory of surgery: operate early in the course of the illness to prevent bad things happening in the future—damage to the kidney or bladder and premature death. Others believed in the quality of life theory: for most men, the natural course of untreated BPH was not life-threatening. For them, the primary goal of surgery was to reduce symptoms and improve well-being.

In a further analysis, we found that mortality rates following surgery in Maine were unexpectedly high. Among men 65 years of age and older, within two months after surgery, 4.3% were dead; within a year, 13.4% were. The increased mortality following surgery was seen in all risk groups, from the healthiest to the sickest.

The evidence that mortality might be higher than previously believed seemed inconsistent with the theory that early surgery increased life expectancy; it focused attention on the need for finding out what the outcomes really were. Thus began a 20-year collaboration between researchers, practicing physicians and academic urologists that led to a revision in the underlying theory for undertaking BPH surgery; established the importance of patient preferences in choice of treatment; championed shared decision-making as strategy for reducing unwarranted variations in surgical treatment and establishing the “right rate”; developed the Patient Outcomes Research Team, or PORT, model for how practicing physicians, researchers, professional leaders, and government can work together to resolve controversy and improve the science of health care delivery through the systematic evaluation of medical innovation; and led to the establishment of the Foundation for Informed Medical Decision-Making (now known as the Informed Medical Decisions Foundation), whose mission is to promote shared decision-making and develop evidence-based patient decision aids to promote informed patient choice.

While many researchers have contributed to the work described here, Drs. Jack Fowler, Al Mulley, and Michael Barry are key contributors to the PORT team concept and our understanding of the importance of patient engagement in decision-making as remedy for unwarranted variation in preference-sensitive care. And each has played a crucial role in the Foundation: Jack Fowler and Michael Barry as its President; and Al Mulley as Founder.



The Preventive Versus the Quality of Life Theory of Surgery for BPH

We undertook a series of studies to clarify the theoretical basis for undertaking surgery, looking first at the preventive theory: Does early intervention improve life expectancy? Drs. Michael Barry and Albert Mulley built an actuarial model predicting life expectancy for those men undergoing surgery and for those choosing what we called “watchful waiting”—expectant treatment that delays surgery until BPH complications or intolerable symptoms appear. While no randomized clinical trial had been reported to provide the probability estimates for the critical events in the model, we found a few cohort studies in the literature that provided estimates of the crossover rates from watchful waiting to surgery; we also used large claims databases to learn about the failure rates among men operated upon (measured by deaths associated with operations and re-operations). The results were reported in:

- Barry MJ, Mulley AG, Jr., Fowler FJ, Wennberg JW. [Watchful waiting vs immediate transurethral resection for symptomatic prostatism. The importance of patients' preferences.](#) *JAMA*. May 27 1988;259(20):3010-7.

The model revealed that for all age groups, early surgery failed to provide a life expectancy gain: The rate of progression of untreated BPH to irreversible end-stage renal disease or life-threatening infection was not sufficient to warrant early intervention with the rationale of improving the chances for survival. But when adjustments were made for possible gains in the quality of life, surgery might prove useful for symptomatic men. However, the opportunity for benefit depended very much on individual patient’s own evaluation of the disutility caused by his symptoms. The conclusion: patient preference should be the dominant factor determining choice.



What are the Gains in Quality of Life?

When we turned to this question, we found the published literature was even less helpful in clarifying what works. Never mind that no randomized clinical trials existed. We couldn't find a cohort study or even a case series report that adequately addressed the subjective responses of patients to surgery. The case for the effectiveness of prostate surgery in improving symptoms was built on anecdote and the analogy that since prostatectomy demonstrably improved urinary flow, it must also improve symptoms and quality of life. To solve this missing data problem, our urologist colleagues agreed to enter their own patients in a prospective study to determine the association between surgery and improvement in the quality of life.

The first task was to learn what mattered to patients and to develop a measurement tool. Dr. Floyd J. Fowler undertook unstructured interviews with patients (using a focus group format), some of whom were contemplating surgery, some of whom were confirmed watchful waiters, and others who had surgery with both good and bad outcomes. From these conversations, he amassed a list of all the outcomes that mattered to patients and then set out to build and validate a questionnaire to quantify symptoms, complications, and quality of life states. He also measured the subjective impact that the BPH condition had on the individual patient—how much he was bothered by his symptoms and what his expectations were from treatment.

When the measurement tool was ready, our urologist colleagues in Maine undertook a cohort study of consecutive patients who underwent prostatectomy. All in all, some 400 patients were interviewed at the time of operation and their outcomes measured three, six, and twelve months after surgery. The results are reported in:

- Fowler FJ, Jr., Wennberg JE, Timothy RP, Barry MJ, Mulley AG, Jr., Hanley D. [Symptom status and quality of life following prostatectomy](#). *JAMA*. May 27 1988;259(20):3018-22.

Just as our urologist colleagues had predicted, the benefits of surgery in reducing symptoms was a “slam bang” effect: about 76% of severely symptomatic men reported only mild or virtually no symptoms after surgery. This outcome was much better than that obtained by watchful waiting, where the best that can be expected is some slight improvement. But the news was not uniformly good. Seventeen percent of those interviewed after surgery reported they still had moderately severe symptoms, and 7% said they were no better off than before. Moreover, well over half reported difficulty with sexual function (retrograde ejaculation), 5% of previously potent men reported no erections, and 4% reported problems with dribbling of urine which they had not had before the operation.

It became clear that the value of prostatectomy as a means for improving the quality of life depended upon the individual patient's own assessment of the risks and benefits. Rational decision-making required the active involvement of the patient in the choice of treatment—in a process we called shared decision-making.



Shared Decision-Making and the BPH Interactive Video Disc

In our study of patients we also learned that men were not all the same in the way they reviewed the impact of BPH upon their lives. Some men, even those with severe symptoms, told us that they were not very much bothered by them. Moreover, as we learned from the interviews conducted prior to surgery, they differed in their concerns about the risk of impotence and incontinence. The more our studies took us into the subjectivity of the patient experience, the more variegated and non-uniform the typology of preference appeared to be. Nothing in the objectivity of the patient—his clinical history, his physical findings, his laboratory scores, his urine flow, even his symptom level—strongly predicted the degree of BPH botheredness or the aversion to the risks of surgery. To learn what the individual patient wants—to make the normatively correct treatment choice from the perspective of the individual—he must be asked to participate in a decision process that disentangles the patient’s preferences from those of others, in particular, the physician.

We began experimenting with ways to promote shared decision-making. Al Mulley, who had explored the concept of interactive videos for other projects, suggested that this medium—a precursor to the internet, with similar interactive features—might be useful for explaining complicated decision problems to patients. The medium proved very useful for explaining the structure of the BPH decision problem: the reality that patients do have a choice was made evident by showing the testimonies of two physician-patients, each severely symptomatic; one who chose surgery, the other watchful waiting. The possible outcomes were made vicariously accessible by filming patients who had good outcomes, as well as those who had experienced the principal complications of surgery and watchful waiting. The computer, which is part of the technology, made it possible to store the probability estimates for the risks and benefits on a subgroup-specific basis. We thus could provide the estimates for the relevant outcomes according to the patient’s age, clinical status, and symptom score.



Learning What Patients Want and Which Rate is Right

The striking differences in attitudes toward their symptoms among clinically similar men with BPH meant that it was not possible for the physician to prescribe the “right” choice of treatment on the basis of any “objective” feature of the physical examination, laboratory test, or even symptom level. The patient needed to reveal his preference by active involvement in the decision process. But would the use of the BPH decision aid lead to better decisions? Could the dilemma of choice—the tradeoff between improvement in urinary tract symptoms and the negative impact of surgery on sexual function—be adequately explained to patients? Would informed patients make choices more consistent with their underlying values? What would be the effect of shared decision-making on the demand for surgery—on the population-based rate of surgery?

In 1995, we published two papers that addressed these questions:

- Barry MJ, Fowler FJ, Jr., Mulley AG, Jr., Henderson JV, Jr., Wennberg JE. [Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplasia](#). *Med Care*. Aug 1995;33(8):771-82.

We learned that patients were able to understand the gist of the decision that needed to be made and that ratings of the program were generally positive at all educational levels. The most important finding: the verification of the importance of the patient’s own evaluation of risks to his sexuality and his attitudes toward symptoms in predicting an individual’s choice of treatment. Separate ratings were developed for symptom level, concern about symptoms, and concern about impotence. When these were used in a multivariate analysis to predict choice, symptom score was not a significant predictor. Despite its superior effect on reducing symptoms (the “main effect” outcome), only one in five men with severe symptoms chose surgery. It became clear that determining the medical necessity of a procedure required more than guidelines based on evidence-based medicine. It required informed patient choice.

- Wagner EH, Barrett P, Barry MJ, Barlow W, Fowler FJ, Jr. [The effect of a shared decision-making program on rates of surgery for benign prostatic hyperplasia. Pilot results](#). *Med Care*. Aug 1995;33(8):765-70.

What would be the demand for surgery—measured as the utilization rate—if shared decision-making were in place? The implementation of the BPH shared decision-making program in two staff model HMOs—the Kaiser Permanente Medical Program in Denver and the Tacoma region of Group Health—gave us the opportunity to evaluate the effect of shared decision-making on population-based rate of BPH surgery. After implementation of shared decision-making, rates dropped about 40% in these populations, compared to controls.



The Open Versus TURP Controversy

The traditional surgical approach to treating BPH is the so-called “open” prostatectomy, performed by making an abdominal incision to gain access to the prostate. By the mid-1980s, the open procedure had been largely replaced by the transurethral prostatectomy (TURP), an operation in which the surgeon reaches the prostate through the urethra. The TURP was widely viewed as safe and effective, more so than the open operation, even though, as is typical of much of surgical innovation, no clinical trials had been undertaken to compare the risks and benefits of the two surgical procedures.

No assessment undertaken by our team produced a more intense debate than the studies showing an increased risk of re-operation and mortality among men treated with TURP when compared to men who had an open prostatectomy. Evidence that TURP might be less safe than assumed first emerged in our 1983 report (unpublished) for the urology study group, which showed a higher than expected mortality rate in the year following surgery (even though many urologists believed the purpose of surgery was to increase life expectancy).

- Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. [Use of claims data systems to evaluate health care outcomes. Mortality and reoperation following prostatectomy](#). *JAMA*. Feb 20 1987;257(7):933-6.

Working together with Noralou Roos, we expanded the study to include claims data for the Canadian province of Manitoba, which showed a similar elevation in mortality, but also documented that patients who underwent the TURP operation were more likely to undergo a second prostate operation compared to those who underwent an open procedure.

- Roos NP, Wennberg JE, Malenka DJ, et al. [Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperplasia](#). *N Engl J Med*. Apr 27 1989;320(17):1120-4.

The replication of the Maine results in Manitoba spurred us on to expand the scope of our investigation to include men undergoing BPH surgery in Maine, Manitoba, Oxford England, and Denmark. In each location and period of time, the cumulative percentage of patients who underwent a second operation was substantially higher for TURPs than for open prostatectomy, as were the long-term age-specific mortality rates. Results could not be explained by age, clinical diagnoses or risk information obtained from anesthesia records.



- Malenka DJ, Roos N, Fisher ES, et al. [Further study of the increased mortality following transurethral prostatectomy: a chart-based analysis.](#) *J Urol.* Aug 1990;144(2 Pt 1):224-7; discussion 228.

We undertook a further study to see if clinical data recorded in the patients' medical records would explain the difference. The addition of chart-based clinical information did not change the result: TURP appeared to be less effective in overcoming urinary obstruction than the open operation and was associated with higher mortality in the years following surgery. However, confounding from still unmeasured patient characteristics could not be ruled out; to resolve this uncertainty, we recommended a randomized clinical trial.

We discussed our recommendation with the leadership of the American Urologic Association, who agreed to organize a clinical trial.



Collaboration with the American Urologic Association

In 1988, the leadership of the American Urologic Association—the professional organization to which most U.S. urologists belong—visited Dartmouth to review our studies. They agreed with our conclusion that, for most men, the goal of surgery was to improve the quality of life—and that it was important for men with BPH to become actively involved in the choice of treatment. This led to an active collaboration to develop standardized tools for assessing symptoms and quality of life for men with BPH. Mike Barry and Jack Fowler took the lead from our group.

- Barry MJ, Fowler FJ, Jr., O'Leary MP, et al. [The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association.](#) *J Urol.* Nov 1992;148(5):1549-57; discussion 1564.

An important result was the AUA Symptom Index—a self-administered questionnaire that quantified symptom level from BPH. Under AUA sponsorship—and after publication in the Association's own journal—it was routinely used by most practicing U.S. urologists, signaling the rapid diffusion of our conclusions that relief of symptoms (not relief of obstruction) was the main reason for surgery in most men.

- Barry MJ, Cockett AT, Holtgrewe HL, McConnell JD, Sihelnik SA, Winfield HN. [Relationship of symptoms of prostatism to commonly used physiological and anatomical measures of the severity of benign prostatic hyperplasia.](#) *J Urol.* Aug 1993;150(2 Pt 1):351-8.

The importance of directly measuring symptom level in assessing the need for surgery was illustrated in a study lead by Mike Barry to evaluate the association between the AUA symptom index and traditional biomedical measures of disease severity in BPH—the mean peak urine flow rate and the size of the prostate gland. Ironically, although the urine flow and prostate size were the diagnosis tests most urologists had been using, they turned out to be uncorrelated with severity of symptoms.



The BPH Treatment Outcomes Study

In the spring of 1989, we met with the AUA to discuss the evidence that the TURP operation (which had been widely adopted into clinical practice without benefit of randomized trials) had worse outcomes than the open prostatectomy, the operation it had replaced. The AUA shared our concern that the association might be causal, and we agreed to work together to organize a clinical trial to test this hypothesis.

Although our initial motivation for the trial was to test the excess mortality hypothesis, we were also determined to put in place the infrastructure and leadership authority to slow the entry of new technologies into practice prior to adequate evaluation. The plan was to build an infrastructure for undertaking sequential trials of all promising innovations in BPH treatment by recruiting leading academic centers as well as physicians in private practice into the network. To test the feasibility of the plan, the AUA invested more than 1 million dollars into a pilot study: BTOPS, the BPH Treatment Outcomes Pilot Study effectiveness trial.

- Cockett AT, Barry MJ, Holtgrewe HL, Sihelnick S, Williams R, McConnell J. [Indications for treatment of benign prostatic hyperplasia. The American Urological Association Study.](#) *Cancer.* Jul 1 1992;70(1 Suppl):280-3.

This article is a report on the design of the BTOPS study and its early success in recruiting patients. (It also demonstrated the commitment on the part of AUA leadership to the project: Dr. Cockett, the principle investigator, was also president-elect of the AUA). The project was envisioned as a twelve-center technology assessment network. We assumed that, with the demonstrated success of the four-center pilot, the full project would easily qualify for federal funding.

BTOPS demonstrated that, with professional leadership, a continuously functioning network of technology assessment centers could be established to compare established technologies to new treatment technologies early in the cycle of innovation. However, our assumption that federal funding would be forthcoming to take the project beyond the pilot phase proved to be wrong. Despite broad agreement that the project would provide critical data for rationalizing treatment choice, clarify whether the TURP operation was associated with higher mortality than the open operation it replaced, and constitute a unique model of professional leadership in the continuous evaluation of new technologies, the project failed to receive the necessary support.



BPH and the PORT Strategy for Patient-Centered Outcomes Research

The first phase of work evolved through collaboration between practicing urologists organized by Dr. Dan Hanley's Maine Medical Assessment Program and our research team, funded by grants from the Commonwealth and the Hartford foundations. Motivated as a response to unwarranted variations in surgery for BPH in Maine (first published in 1975), the collaboration led to clarification of the underlying reasons for surgery and established the importance of patient preference in the choice of treatment and the active involvement of patients in clinical decision-making—in shared decision-making. Some thirteen years later, we published an overview of the research methods and strategies, as well as a summary of the principle findings:

- Wennberg JE, Mulley AG, Jr., Hanley D, et al. [An assessment of prostatectomy for benign urinary tract obstruction. Geographic variations and the evaluation of medical care outcomes.](#) *JAMA*. May 27 1988;259(20):3027-30.

In September 1989, our research team became part of a new federal outcomes research program—the Prostate Disease Patient Outcomes Research Team. Modeled to a large extent after our Maine project, the PORT program, under the auspices of the Agency for Health Care Policy and Research (AHCPR), offered five-year grants to teams that agreed to pursue a course of discovery to uncover differences in medical opinion and undertake clarifying research to test clinical hypotheses. In the fall of 1994, I was asked to give a talk at a festschrift for Ian Chalmers and Richard Peto and took the opportunity to provide an update on our research strategy for BPH (including the BTOPS effectiveness trial) and describe our thinking concerning ways to design clinical trials that would allow the evaluation of the impact of patient preferences on clinical outcomes:

- Wennberg JE, Barry MJ, Fowler FJ, Mulley A. [Outcomes Research, PORTs, and Health Care Reform.](#) *Ann NY Acad Sci*. 1993;703:52-62.
- Wennberg JE. [On the status of the Prostate Disease Assessment Team.](#) *Health Serv Res*. Dec 1990;25(5):709-16.



The Prostate Port and Early Stage Prostate Cancer

In the early 1990s, our PORT team began to apply its methods to prostate cancer. Here are five papers that illustrate different aspect of our approach to outcomes research.

Critical assessment of published literature:

- Wasson JH, Cushman CC, Bruskewitz RC, Littenberg B, Mulley AG, Jr., Wennberg JE. [A structured literature review of treatment for localized prostate cancer. Prostate Disease Patient Outcome Research Team. Arch Fam Med. May 1993;2\(5\):487-93.](#)

An early step in the evaluation of prostate cancer treatment was to look at the effectiveness of treatment. This paper reports the findings of a critical review of the literature, led by John Wasson, to evaluate the clinical course of localized prostate cancer, treated by radical prostatectomy, radiation therapy or watchful waiting. The results were disappointing: we were unable to determine treatment effectiveness because the quality of published articles was simply inadequate. We did, however, find evidence that surgery and radiation were associated with significant complications, including incontinence, impotence and urethral stricture.

Our conclusion: “Until better scientific evidence is available, patients and their physicians cannot make informed choices based on knowledge of the benefits of radical prostatectomy, radiation, or watchful waiting.”

Use of claims data:

- Lu-Yao GL, McLerran D, Wasson J, Wennberg JE. [An assessment of radical prostatectomy. Time trends, geographic variation, and outcomes. The Prostate Patient Outcomes Research Team. JAMA. May 26 1993;269\(20\):2633-6.](#)

This 1993 study headed by Grace Lu-Yao served as an early warning about an emerging epidemic in prostate cancer surgery: over a six-year period, the number of radical prostatectomies performed increased almost eightfold, from about 2,000 procedures in 1984 to 16,000 in 1990. Striking geographic variation was documented—from 20 procedures per 100,000 Medicare men in Rhode Island to more than 400 in Alaska. All states in New England and the mid-Atlantic region had rates below 60 per 100,000, while all states in the Pacific and Intermountain region had rates in excess of 130 per 100,000. The paper also reported on complications following prostate cancer. For example, for men 75 years of age and older, “almost 2% died and nearly 8% suffered major cardiopulmonary complications within 30 days of the operation.”

Our conclusion: “The sharp increase and wide geographic variation in radical prostatectomy rates make the evaluation of this surgical practice a pressing issue.”



Patient interviews:

- Fowler FJ, Jr., Barry MJ, Lu-Yao G, Roman A, Wasson J, Wennberg JE. [Patient-reported complications and follow-up treatment after radical prostatectomy. The National Medicare Experience: 1988-1990](#) (updated June 1993). *Urology*. Dec 1993;42(6):622-9.

Gaps in critical information on patient experience, quality of life and degree of bother can be filled by interviewing patients before and after intervention. Jack Fowler led our effort to improve the information on complications for prostate cancer surgery. Working together with the Medicare program, we arranged to interview a national sample of Medicare patients who, according to their claims history, had undergone a radical prostatectomy two or more years previously. Our results were thus representative of the U.S. population enrolled in traditional Medicare. Over 30% of patients who underwent this operation reported they were incontinent of urine—to the point where they needed to wear pads or use clamps—only 11% reported that they had any erection sufficient for intercourse, and 20% reported having treatment for urethral strictures after their cancer surgery.

Our conclusion: “These estimates of complications and follow-up treatment rates are generally higher ... than estimates previously published. Patients and physicians may want to weigh heavily the complications and need for follow-up treatments when considering radical prostatectomy for prostate cancer.”

Decision analysis:

- Fleming C, Wasson JH, Albertsen PC, Barry MJ, Wennberg JE. [A decision analysis of alternative treatment strategies for clinically localized prostate cancer. Prostate Patient Outcomes Research Team](#). *JAMA*. May 26 1993;269(20):2650-8.

As we did for BPH, we conducted a decision analysis to evaluate the impact on men with localized prostate cancer according to the treatment option chosen: radical prostatectomy, radiation therapy, and watchful waiting. Craig Fleming led our team, using a Markov model and the probabilities and rates for important clinical events that came from our analysis of claims data and the review of the literature. The model legitimized watchful waiting as a treatment option; it predicted that surgery or radiation may benefit selected groups of patients, but for most, the potential benefits are small enough that choice of treatment is sensitive to patient preferences, particularly concerns about complications.

Our conclusion: “The choice of watchful waiting is a reasonable alternative to invasive treatment for many men with localized prostatic carcinoma.”



Patient decision aids:

- Flood AB, Wennberg JE, Nease RF, Jr., Fowler FJ, Jr., Ding J, Hynes LM. [The importance of patient preference in the decision to screen for prostate cancer. Prostate Patient Outcomes Research Team. J Gen Intern Med. Jun 1996;11\(6\):342-9.](#)

The choice of treatment for early stage prostate cancer turned out to be a classic example of decision-making under scientific uncertainty about the effect of treatment on the main outcome—a reduction in mortality rates, but with substantial information concerning risks of treatment options and evidence that treatment choice varies according to patient preferences. Because the epidemic in prostate cancer surgery was linked to increased use of the prostate-specific antigen (PSA) test (to screen for prostate cancer), we developed a decision aid to help men decide if they wanted to be screened. Key to informed choice about screening is an understanding of what is known and not known about the consequences of treatment, should prostate cancer be found.

We tested the effect of the PSA decision aid on clinic patients and on patients attending a public event advocating PSA use. In this study, led by Ann Flood, patients who used the PSA decision aid were much more knowledgeable than controls about the uncertainty concerning the efficacy of treatment, about the natural history of watchful waiting, and about the predictive value of the test. Moreover, among clinic patients, they were much more likely than the controls to prefer watchful waiting (86% versus 40%), while the control group was much more likely to want surgery (41% versus 8%). Actual use of PSA testing was significantly less—only 12% had a PSA test at their next physician visit, compared to 23% of the controls.

Our conclusion: “Preference regarding cancer screening and treatment is greatly influenced by information about medical uncertainties. Because informed patient choices vary, PSA screening decisions should incorporate individual [patient] preferences.”



PORTS and Politics: The Rise and Fall of Medical Effectiveness Research

During the late 1980s, with health care reform at the top of congressional priority, the problem of practice variation was discussed by key members of the U.S. Congress, in particular George Mitchell and David Durenberger, who championed legislation to establish research teams to address practice variations using the collaborative Maine Medical Assessment program as a model. This led to legislation to fund Patient Outcomes Research Teams and, a year later, a new agency—the Agency for Health Care Policy and Research—was founded and given responsibility for organizing a broad outcomes research agenda, including the PORT program.

The rise of the PORT program and its origins in Maine as a response to practice variation was profiled in a 1992 article by Bradford Gray in *Health Affairs*.

- Gray BH. [The legislative battle over health services research](#). *Health Aff (Millwood)*. Winter 1992;11(4):38-66.

The first sign of trouble was the failure to find federal funding for the clinical trial network organized by the AUA and our PORT to test the open versus TURP controversy and to set up an infrastructure for the sequential evaluation of new technologies for treating BPH against established treatments. Despite the success of BTOPS in demonstrating the feasibility of such a network, neither the NIH nor the Agency for Health Care Policy and Research (AHCPR) would agree to fund the project; in February 1993, the AUA and the PORT gave up the follow-up of the 170 patients already randomized as part of BTOPS and abandoned the concept of setting up a technology assessment network after four years, three rejected grants, and over \$1 million of BTOPS funding from AUA membership dues.

With the change in Congress in 1994, congressional support for AHCPR took a nosedive; the PORT strategy was abandoned, in part because of lobbying efforts on the part of dissident back surgeons whose favorite surgical approach to back pain had been challenged by the back pain PORT, and in part because of distrust of AHCPR for its alleged role in the Clinton health plan proposal. Again, Brad Gray and his colleagues monitored this change, with a 2003 article in *Health Affairs*.

- Gray BH, Gusmano MK, Collins SR. [AHCPR and the changing politics of health services research](#). *Health Aff (Millwood)*. Jan-Jun 2003;Suppl Web Exclusives:W3-283-307.

For additional commentary on the rise and fall of medical effectiveness, see:

- Wennberg JE. [AHCPR and the strategy for health care reform](#). *Health Aff (Millwood)*. Winter 1992;11(4):67-71.
- Wennberg JE. [The more things change...: the federal government's role in the evaluative sciences](#). *Health Aff (Millwood)*. Jan-Jun 2003;Suppl Web Exclusives:W3-308-10.
- Deyo RA, Psaty BM, Simon G, Wagner EH, Omenn GS. [The messenger under attack--intimidation of researchers by special-interest groups](#). *N Engl J Med*. Apr 17 1997;336(16):1176-80.



The Foundation for Informed Medical Decision-Making

Under the assumption that the Agency for Health Care Policy and Research's PORT program would provide continuous evaluation of new as well as established treatment options—and that choice among these options should be based on patient preferences—we anticipated the need for an organization that whose job was to build and update patient decision aids along the lines developed by the Prostate PORT team. But what should such an organization look like?

First, it could not be an organization with vested interests in any one of the treatment options available. It couldn't be sponsored by a pharmaceutical or medical device company, a medical specialty group such as the American Urologic Association, an insurance company, or anyone else that stood to lose or gain from a change in utilization rates such as we had seen following the introduction of the shared decision-making video for BPH. In theory, the federal government could have provided funding, but there were no provisions in federal science policy for this task. Our solution was to create the [Foundation for Informed Medical Decision-Making](#), a not-for-profit educational and research organization tasked with this responsibility.

Although there have been many changes in federal policy and private charity funding, the Foundation has adapted to these changes and remains an important source of unbiased information for patients concerning their treatment options. On March 31, 2014, the Foundation and Healthwise merged to form a single organization.



The Epidemiology of Medical Care

These papers describe research into the epidemiology of medical care, conducted from early 1980s until the present day. The common themes are the description of medical practice variation; the causes of variation; and the evaluation of the outcomes and efficiency of health care.

1. [How Much Variation is There?](#)

- Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison
- Will Payment Based on Diagnosis-Related Groups Control Hospital Costs?
- Using Diagnosis-Related Groups for Studying Variations in Hospital Admissions

2. [The Boston New Haven Trilogy](#)

- Are Hospital Services Rationed in New Haven or Over-used in Boston?
- Hospital Use and Mortality Among Medicare Beneficiaries in Boston and New Haven
- Hospital Re-admission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven

3. [The Dartmouth Atlas of Health Care](#)

4. [Is More Better?](#)

- The Implications of Regional Variations in Medicare Spending: Part 1. Utilization of Services and the Quality of Care
- The Implications of Regional Variations in Medicare Spending: Part 2. Health Outcomes and Satisfaction with Care
- Variations in Longitudinal Efficiency of Academic Medical Centers
- Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care
- Hospital Quality and Intensity of Spending: Is There an Association?
- Relationship between Regional Per Capita Medicare Expenditures and Patient Perceptions of Quality of Care
- Inpatient Care Intensity and Patients' Ratings of Their Hospital Experiences

5. [Race, Poverty, and Utilization](#)

- Who You Are and Where You Live: How Race and Geography Affect the Treatment of Medicare Beneficiaries
- Racial, Ethnic, and Geographic Disparities in Rates of Knee Arthroplasty among Medicare Patients



6. [Evaluating Provider Efficiency in Managing Chronic Illness](#)

- Use of Hospitals, Physician Visits and Hospice Care among Cohorts Loyal to Highly Respected Hospitals in the United States
- Use of Medicare Claims Data to Monitor Provider-Specific Performance Among Patients with Severe Chronic Illness
- Evaluating the Efficiency of California Providers in Caring for Patients with Chronic Illness

7. [Recent Research: Risk Adjustment and the Observational Intensity Bias](#)

- Regional Variations in Diagnostic Practices
- Geographic Variation in Diagnosis Frequency and Risk of Death among Medicare Beneficiaries
- Observational Intensity Bias Associated with Illness Adjustment: Cross Sectional Analysis of Insurance Claims
- A Population Health Approach to Reducing Observational Intensity Bias in Health Risk Adjustment: Cross Sectional Analysis of Insurance Claims.



How Much Variation is There?

In the mid-1970s, Klim McPherson and I worked together at the Center for the Analysis of Health Practice at Harvard's School of Public Health. Our collaboration led to two papers that substantially increased our understanding of the variation phenomenon. First, it became clear that surgical practice varied as much in the U.K.'s nationalized, centrally-managed health care system as it did in New England's fee-for-service system. Second, variation in use of hospitals appears to be ubiquitous, affecting most causes of admission for medical conditions as well as for surgical conditions.

- McPherson K, Wennberg JE, Hovind OB, Clifford P. [Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway](#). *N Engl J Med*. Nov 18 1982;307(21):1310-4.

To address the question of whether regional rates for common surgical procedures are more uniform in nationally-organized systems of care than in the U.S, we compared the incidence of surgery in counties in southern Norway, in districts in the West Midlands of the United Kingdom, and in Maine, Rhode Island, and Vermont.

We found that, for most procedures, the pattern of variation was similar across international boundaries. Although rates for procedures such as tonsillectomy, hysterectomy, and prostatectomy were lower in Norway and the U.K., the variations among regions were as great as in the United States. We interpreted the similarity in practice patterns as evidence against the notion that clinical decision-making in the low-rate countries reflects a clear threshold above which everyone receives the treatment and below which decisions are carefully balanced on the basis of competing needs and priorities. Small area variation appeared to be an international phenomenon, independent of national strategies for organizing and financing medical care.

In order to make the comparisons in this paper, we first needed to solve a methodological issue: how to compare variation among regions and procedures when the size of the populations and the incidence of various surgeries varied from place to place and procedure to procedure. Faced with this challenge, Klim McPherson and Peter Clifford developed a statistical measure of variation that is independent of scale: the Systematic Coefficient of Variation, or SCV. The SCV has played an important role in our research to rank surgical procedures and medical causes of hospitalization according to degree of variation.

- Wennberg JE, McPherson K, Caper P. [Will payment based on diagnosis-related groups control hospital costs?](#) *N Engl J Med*. Aug 2 1984;311(5):295-300.

How much variation is there in the use of hospitals according to the causes of admission? Our early papers were restricted to a few common procedures or large groups of medical conditions such as "respiratory diseases," which masked variation for specific conditions such as pneumonia and chronic obstructive pulmonary disease. However, with the introduction of the diagnosis-related groups (DRG) system for classifying hospitalized patients, and the development of the SCV as a tool for ranking relative variation, we were able to evaluate the degree of variation for the entire hospitalized case mix.



This study looked at variation in hospitalization rates for DRGs among hospital service areas in Maine from 1980-1982. An important feature of the analysis was the classification of hospitalizations into categories according to relative variation—low, moderate, and high-variation. The boundary between categories was scaled empirically according to the SCV for common surgical procedures. We picked the SCV for hysterectomy—widely documented as a highly variable discretionary procedure—to serve as the lower boundary for the high-variation category. By this standard, 90% of DRGs fell into the high-variation category, suggesting that physician discretion plays an important role in influencing utilization for most surgical procedures and medical causes of hospitalization.

The study was formulated as a critique of the prospects for the DRG system to reduce costs. The DRG reimbursement system focused on containing cost per episode of hospitalization, presumably under the assumption that the admission rate was constrained by professional guidelines. Building on the findings reported in [Maine III](#), our study demonstrated that the admission rate was more important than length of stay in explaining variation in use of hospital beds (the admission rate explained 69% of the variation in hospital bed use; length of stay only 15%). Unless a shortening of length of stay (induced by DRG payment policy) was coupled with a reduction in capacity, admission rates would likely go up, with the net effect on the DRG program being an increase in per capita costs for hospitals.

- Roos NP, Wennberg JE, McPherson K. [Using diagnosis-related groups for studying variations in hospital admissions](#). *Health Care Financ Rev.* Summer 1988;9(4):53-62.

We subsequently repeated the DRG-based analysis in Iowa, California, New York, and Massachusetts, with similar results. Less than 10% of hospitalizations were for DRG conditions that were classified as low or moderate variation (i.e., less variable on the SCV scale than hysterectomy). Moreover, the DRG groups were consistent in their pattern of variation among the states. Several communities contained well-known academic medical centers, including the University of Iowa, Stanford, the University of Rochester, the University of Massachusetts, the University of California (Sacramento-Davis), Yale, and the three medical schools located in Boston. It was becoming clear that the practice variation phenomenon was ubiquitous in medically “advanced” communities served by those academic medical centers, as well as in rural Vermont, Maine, and Iowa. Among the causes of hospitalization, high variation was the rule, low variation the exception.



The Boston/New Haven Trilogy

The Boston-New Haven comparisons probably did more to convince the policy community and leaders in medicine that the practice variation phenomenon is a central feature of the health care market than any other small area study we undertook. Previous studies could be discounted on the basis that Maine and Vermont did not reflect what was going on in high-quality places. As the site of some of world's most prestigious medical centers, widely recognized by the public and profession alike as places for high-quality care and scientific excellence, Boston and New Haven were the perfect examples: communities that were accessible to the experience of leading policymakers and clinicians throughout the country, many of whom had been educated or trained there. How could it be that the highest quality care produced for Bostonians could require twice as many resources as the highest quality care for New Havenites?

The first inkling of the magnitude of the difference came from the data shown in Exhibit 4 of a 1978 study of hospital resource allocation, "[Dealing with Medical Practice Variations: A Proposal for Action.](#)" In that year, the outlay for hospital care for Bostonians was twice that for New Havenites, \$448 per person compared to \$215; the hospital industry invested 18.2 per 1,000 hospital employees in the health of Bostonians, compared to 9.5 for New Haven residents; residents of Boston used 4.4 hospital beds per 1,000 compared to only 2.7 for New Havenites. In each community, the large majority of hospital care was rendered by hospitals affiliated with well-known medical schools: Yale, Harvard, Tufts, and Boston University.

The 1980s brought the opportunity to turn the spotlight of small area variations more fully on the heartland of scientific medicine. We acquired the needed data to study the details of variations in the hospitalized case mix between these two communities. The results were the three papers that follow. The first paper raised the "which rate is right" question—whether health care was rationed in New Haven or overutilized in Boston. The second, which used Medicare data, asked questions about the outcomes of care: among these demographically similar cities with a twofold difference in Medicare outlays for hospital care, were mortality rates different? The third adopted a new approach to characterizing variations, showing that a cohort approach to the analysis of Medicare data gives similar results as classic small area analysis, but is also useful for studying the patterns of care for individual hospitals and the long-term outcomes of care.

- Wennberg JE, Freeman JL, Culp WJ. [Are hospital services rationed in New Haven or over-utilised in Boston?](#) *Lancet*. May 23 1987;1(8543):1185-9.

With completion of the Maine DRG study, we had a method for characterizing in detail the hospitalized case mix of defined populations. This article was the first to use the utilization experience of a low-rate area as a benchmark to characterize the case mix of a high rate area. How many "extra" beds are used in the care of Bostonians, compared to New Havenites? For what clinical purposes are the extra beds of Boston used? The approach uses the classic epidemiologic strategy of estimating relative risk: for what categories of cases are the risks of hospitalization higher for Bostonians? It also employed the less often used estimate of attributable or "absolute" risk: among the various categories of cases, how much of the excess beds does each account for?



As the article shows, most of the excess beds of Boston were invested in high-variation medical conditions among patients with chronic illnesses. Hospitalization rates for minor surgery were affected, but while the overall rate for major surgery was very nearly the same in both communities, individual procedures varied idiosyncratically, some higher in Boston, others higher in New Haven (just as we had seen in early studies in Maine).

Curious to learn whether clinicians in New Haven were aware that resources and utilization for medical (mostly chronic) conditions were much lower, or that surgery varied the way it did, I met with the chiefs of the medical and surgical staffs at Yale-New Haven Hospital. They were unaware of the magnitude of the difference; indeed, even though many had also been trained or practiced in Boston, until I showed them the data, there was no consensus concerning which community might have the higher rates. These conversations made it clear to me that there was no sense of lack of resources or fear of health care rationing among New Haven physicians, even though from a regional as well as a national perspective, the allocation of resources to New Havenites was very modest. The occupancy rate of the hospitals in New Haven, which at that time was relatively low and about the same for Boston teaching hospitals, stood as additional evidence against rationing: beds were available if needed.

- Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. [Hospital use and mortality among Medicare beneficiaries in Boston and New Haven](#). *N Engl J Med*. Oct 26 1989;321(17):1168-73.

This second Boston/New Haven paper took up the question of health care outcomes: was the lower investment in acute hospital care among New Haven residents associated with higher mortality, compared to the demographically similar residents of Boston? The answer was that there was no discernible difference. Age, race, and sex-standardized population mortality was virtually the same in both communities. We concluded that the systematic study of the outcomes of high variation conditions should have the highest priority on the research agenda “(b)ecause the answers may create opportunities to reduce the cost of medical care without damaging the welfare of patients.”

The paper makes an important point about limitations on the use of Medicare data in evaluating the quality of care. At the time of its publication, there was considerable enthusiasm in some research circles for interpreting “severity-adjusted” case fatality rates for medical conditions, such as congestive heart failure, as an indicator of the quality of hospital care. To avoid the bias associated with varying length of stay in hospital, these methods typically called for measuring mortality up to 30 days after admission. When we applied this measure to evaluate the “quality of care” between Boston and New Haven, we found case fatality rates for high-variation conditions were considerably higher in New Haven. However, as the paper points out, “these statistics did not imply differences in the skill of clinicians and hospitals in Boston and New Haven in preventing unnecessary deaths, but resulted instead from differences in clinical policies.” The differences were explained by differences in clinical policies governing admissions and readmission, length of stay, and place of treatment of the terminally ill between the two cities. “These factors differ simultaneously and variably from region to region, between communities within regions, and from hospital to hospital within communities ...”

Our conclusion: “(V)ariations in hospital-based mortality rates among patients with high-variation medical conditions should not be interpreted as reflecting differences in clinical skill or productivity.”



- Fisher ES, Wennberg JE, Stukel TA, Sharp SM. [Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven](#). *N Engl J Med*. Oct 13 1994;331(15):989-95.

Prior to this report, small area analyses had depended upon the population-based method of classic epidemiology (in which the numerator for rates is an event and the denominator the resident population). This paper demonstrated that utilization differences seen between Boston and New Haven at the population level were replicated in the subsequent readmission rates among cohorts of patients defined by a hospitalization for a “low-variation” hospitalization such as hip fracture. The study shows that over a three-year follow-up after discharge for such a condition, residents of Boston experienced about a 1.6 higher relative risk in each six-month period. The cohort study made it possible to test hypotheses concerning the association between capacity, utilization, and outcomes of care in models in which the question of initial differences in severity of illness among areas (or hospitals) was narrowly constrained to the severity of a specific condition such as a hip fracture or acute myocardial infarction.

The paper shows this method can also be used to ascertain the relative risk for readmission according to the hospital the patient uses. I believe this relative risk identifies the effect of the (otherwise unmeasured) per capita bed supply on the (otherwise unmeasured) population-at-risk of the individual hospital; the ratio is thus a measure of the tendency toward supplier-induced demand as it operates at the level of the individual hospital, and in subsequent work proved useful in evaluating hospital performance in urban areas.

The study added a new dimension to the debate about whether the readmission rates were determined by patient illness. In order for this argument to hold, the data in Table 1 of the paper required that myocardial infarction, stroke, hip fracture, and cancer patients admitted to the Yale-New Haven Hospital be the least ill, and that patients with the same diagnosis admitted to the Massachusetts General Hospital and Boston City Hospital were less sick than those admitted to Boston University Hospital, St. Elizabeth’s Hospital, and Carney Hospital. No clinician in either Boston or New Haven stepped forward to defend this interpretation. Indeed, clinicians at individual hospitals found it impossible to identify their own institutions in Table 1, but none felt that patients at the Massachusetts General Hospital or Boston City Hospital were less ill than those at University Hospital, which had a higher readmission ratio. Ironically, the staff physicians and house staff who served the patients loyal to Boston City Hospital (which serves the poor, and therefore presumably the more severely ill) were the same physicians who served the patients who received care at University Hospital. They were simply unaware that they practiced medicine differently, depending on which medical service they were on.



The Dartmouth Atlas of Health Care

The outcome of the presidential election of 1992 meant to many that health care reform was finally going to happen and the federal government would play a central role in shaping it. Whether reform emphasized regulation, as in the Canadian approach, or competition, as in the “managed competition” model proposed by the Jackson Hole Group, we at Dartmouth felt there would be new demands for accurate information describing the performance of the health care system on a local and regional as well as a state and national basis; and there would be need for objective reports describing the impact of health reform on the American people. Thanks to a generous grant from the Robert Wood Johnson Foundation, we undertook to develop a uniform set of population-based performance indicators based on small area analysis and planned to monitor changes in the delivery system that would occur as federal health reform progressed. We also undertook research to improve the measures of resource allocation and study the relationship between capacity, utilization, and outcomes.

As the shape of the Clinton health plan proceeded, we geared our work to anticipate the informational needs of the Alliances that were intended to manage the financing of care on a regional basis. We conducted a series of patient origin studies to define hospital service areas and hospital referral regions throughout the United States and, for these areas and regions, estimated population-based resource inputs (dollars, workforce and facilities) and utilization of medical and surgical care for each area and region. We planned the initial release of the Atlas to be timed to passage of the legislation. By late summer 1994, however, it became apparent that health reform would not follow a uniform federal plan; indeed, “reform” was already occurring and would accelerate because of the actions of employers and competing health care plans. If regulation were to play any role, it would be primarily because of what state governments, rather than the federal government, would do.

While this spectacular shift necessarily influenced our thinking about the target audiences for the Atlas, it did not alter our conviction on the need for the feedback to the American people of accurate information on the status of health care markets, including the changes that would occur in the delivery system as a result of private sector initiatives. Working together with the Robert Wood Johnson Foundation, the Health Care Finance Administration, and the American Hospital Association, we put in place the infrastructure to undertake the research and development essential to [The Dartmouth Atlas of Health Care](#), the first volumes of which appeared in 1996.



Is More Better?

The uncovering of the striking variation in health care delivery naturally focused our attention on the question of outcomes. Much of our research involved the evaluation of specific treatment hypotheses for well-defined conditions, such as alternative ways of treating benign prostatic hyperplasia or prostate cancer. Work in this area required a number of different methodologies and approaches, as well as an emphasis on the role of patient preference in the choice among treatment options. Indeed, for preference-sensitive care, the “which rate is right?” question cannot be addressed in the absence of informed patient choice.

The question of whether greater aggregate spending for health care leads to better health care outcomes presents a different challenge. Greater per capita Medicare spending among regions is *not* associated with higher rates of elective surgery or effective care such as life-saving drugs following heart attacks. It *is* closely associated with greater intensity of care: more physician visits, hospitalizations, stays in ICUs, diagnostic tests, and imaging exams per capita, mostly for patients with chronic illnesses. The variation in care intensity occurs without clear hypotheses as to why more is better or even awareness on the part of local physicians as to where their region stands on the spectrum of variation in care intensity.

[The Dartmouth Atlas project](#) provided an opportunity to examine the “more is better” hypothesis, using data for the entire United States. In a series of studies, researchers examined the association between the intensity of care and the quality and outcomes of care, including the patient experience. I summarized the major findings in *Tracking Medicine*. The bottom line: at the levels of resource inputs seen in this country, greater spending and care intensity does not appear to result in higher technical quality of care, lower mortality rates, or an improved patient experience.

- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. [The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care.](#) *Ann Intern Med.* Feb 18 2003;138(4):273-87.

This study, headed by Elliott Fisher, examined the association between intensity of care (measured by spending during last six months of life) and survival and patient satisfaction. It used a cohort design: a five-year follow-up of patients who had a heart attack, an admission for colon cancer or a hip fracture, plus a random sample of Medicare enrollees who were part the Medicare Current Beneficiary Survey (MCBS). The cohorts were studied according to the region where they lived.

Part 1 characterized the differences in patterns of medical practice according to variation in the overall intensity of care—defined as the level of spending in the patient’s hospital referral region of residence during the last six months of life. Increasing care intensity was associated with more physician visits, diagnostic procedures, imaging exams, hospital admissions, and days spent in intensive care units. By contrast, surgical procedure rates were unassociated with overall care intensity: rates of gall bladder surgery, cataract surgery, hip and knee replacement, and back surgery were the same in the regions that ranked at the bottom of the care intensity index as in those at the top. Finally, the quality of care, measured as the percentage of the population getting effective care (such as a beta-blocker after a heart attack or a flu shot), was actually better in regions that ranked at the bottom in intensity of care.



- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. [The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care.](#) *Ann Intern Med.* Feb 18 2003;138(4):288-98.

Part 2 looked at mortality for up to five years following the initial hospitalization for a hip fracture, heart attack, or colon cancer. More did not prove to be better; indeed, each 10% increase in care intensity was associated with an increase in relative risk of death: 1.003 for hip fracture; 1.007 for heart attack; and 1.012 for colon cancer. The findings were consistent across age groups, gender, race, income, and characteristics of the hospital. The study also showed no association between care intensity and changes in functional status and patient satisfaction.

The authors' conclusion was quite conservative:

“Medicare enrollees in higher-spending regions receive more care than those in lower-spending regions but do not have better health outcomes or satisfaction with care.”

- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ. [Variations in the longitudinal efficiency of academic medical centers.](#) *Health Aff (Millwood).* 2004;Suppl Variation:Var19-32.

In this 2004 paper in Health Affairs, Elliott Fisher and his colleagues replicated the *Annals* papers, this time concentrating on hip fracture, heart attack, and cancer patients who were initially hospitalized in an academic medical center associated with a medical school. Again, greater care intensity was not associated with improved life expectancy (nor better quality).

- Baicker K, Chandra A. [Medicare spending, the physician workforce, and beneficiaries' quality of care.](#) *Health Aff (Millwood).* Jan-Jun 2004;Suppl Web Exclusives:W4-184-97.

This paper examined the association between Medicare spending and the use of effective care. Do patients in regions with greater Medicare spending receive better care as measured by objective process quality measures such as percent of heart attack patients who receive a necessary, life-saving medication? The authors found the opposite: higher levels of spending were associated with worse quality scores. Moreover, regions with a predominance of medical specialists as compared to primary care physicians tended to have worse scores.



- Yasaitis L, Fisher ES, Skinner JS, Chandra A. [Hospital quality and intensity of spending: is there an association?](#) *Health Aff (Millwood)*. Jul-Aug 2009;28(4):w566-72.

This paper looked further at the association between spending level and quality, this time at the individual hospital level. The results confirmed the previous finding: hospitals with high-intensity patterns of care had worse quality scores than those with low-intensity care.

- Fowler FJ, Jr., Gallagher PM, Anthony DL, Larsen K, Skinner JS. [Relationship between regional per capita Medicare expenditures and patient perceptions of quality of care.](#) *JAMA*. May 28 2008;299(20):2406-12.

This study looked at the association between care intensity and the patient experience. Medicare enrollees were asked to rate their care as to its quality. The researchers uncovered an inverse relationship between Medicare spending and patients' global rating of their care. For example, 63% of respondents living in low-cost regions gave their care a high score, compared with 55% of those in high-cost regions. An inverse relationship was also found between spending and access: 8% of respondents in high-spending regions reported unmet need for specialists care, while in low-spending regions, only 3% did. The bottom line: across the nation, greater care intensity was certainly not associated with greater satisfaction and lowered perception of unmet need and, paradoxically, perceptions of scarcity for specialists care was greater in regions with more specialists on a per capita basis.

- Wennberg JE, Bronner K, Skinner JS, Fisher ES, Goodman DC. [Inpatient care intensity and patients' ratings of their hospital experiences.](#) *Health Aff (Millwood)*. Jan-Feb 2009;28(1):103-12.

This paper looked at the association among regions between intensity of care, measured by the HCI index [1], and the patient's experience while hospitalized. It took advantage of a new CMS survey of recently hospitalized Medicare patients who were asked to provide a global rating of their hospital, similar to that used in the Fowler study. Patients from high-care-intensity regions gave their hospitals lower ratings than those from low-intensity regions. Indeed, patients living in high-intensity regions ranked their hospitals lower on all ten reported aspects of care in the CMS survey, including communication between doctors, nurses and patients, patient pain control, discharge planning, and overall rating.

We also looked at the association between the HCI index, patient rating of hospitals, and objective technical quality measures. Just as predicted from previous studies, regions with greater care intensity tended to have worse quality scores. But we also noticed a puzzling correlation between quality measures and patient ratings: hospitals in regions with lower patient ratings tended to have worse quality scores.



The correlations suggest a causal pathway: chaotic, disorganized care results in less attention to patient needs and wants. At the same time, in high-intensity regions, many more physicians, in particular medical specialists, are involved in the care of the patient, often with no one in charge. This, in turn, leads to substandard care, repeated and unnecessary diagnostic tests and services, and worse performance on technical quality measures. Organized systems of care, prevalent in low-intensity regions, use many fewer resources in managing chronic illness and are able to coordinate services to better achieve therapeutic goals, including those measured in CMS quality scores. (For further discussion, see pages 167-168 in *Tracking Medicine*.)

[1] The HCI index is a summary measure of the intensity of inpatient care. It is based on two supply-sensitive utilization measures: the average number of days patients spent in hospital and the average number of physician visits patients experienced while in the hospital. These are highly correlated with hospital beds, physician supply, and Medicare spending, but are not biased by differences in prices across regions or by the way Medicare pays for its services. The index is computed as the average of two ratios: the ratio of the number of inpatient days in a region or hospital cohort, compared to the national average, and the ratio of the number of inpatient physician visits per patient, also compared to the national average. The HCI index can be calculated for any cohort of patients and for any fixed interval of time, such as for cohorts of chronically ill patients during six-month intervals prior to death and for the last two years of life.



Race, Poverty, and Utilization

An extensive literature documents the association between race, poverty, and utilization rates, with most studies showing lower use among disadvantaged populations. These differences are commonly interpreted to mean underuse, as “disparities” that signal failure to obtain access to necessary care. As the Dartmouth Atlas project matured, we accumulated enough data to study the effects of race and poverty on the patterns of practice among regions and hospitals—and to interpret these effects in terms of the category of care: effective care, preference-sensitive care, and supply-sensitive care.

- Baicker K, Chandra A, Skinner JS, Wennberg JE. [Who you are and where you live: how race and geography affect the treatment of Medicare beneficiaries](#). *Health Aff (Millwood)*. 2004;Suppl Variation:Var33-44.

This paper illustrates the importance of understanding geographic variation as well as the category of care in reaching an interpretation of disparity (variation) in utilization rates. Most studies are based on national samples that commonly show lower average rates for minority groups than for the white population. This study, which examines black-white differences in utilization, shows that the overall national average masks considerable variation in utilization among regions for whites as well as blacks; moreover, for the supply-sensitive category, the disparity is in the opposite direction: if judgment is based on the average national rate for end-of-life intensity of care, whites are underusing care compared to blacks.

By illustrating that no “simple story” explains regional variation, the study emphasizes the importance of keeping the categories of care clear in interpreting variation in treatment rates and in proposing solutions. For effective category, underuse affects both whites and blacks, variably so from one region to another. Here, the proper remedy is to assure access to everyone in need. For preference-sensitive procedures such as back surgery, black and white rates vary extensively, but the “right rate” isn’t clear, and an important reason is that patients—black and white—are not actively involved in treatment decisions. Here, the proper remedy is to engage the patient in the choice of treatment in ways that assure the individual patient’s preferences are respected. For supply-sensitive care, the question is relative efficiency in managing care over time, particularly for those with chronic illness. Here, the proper remedy is organized care, surely not increasing the intensity of acute hospital care.

- Skinner J, Weinstein JN, Sporer SM, Wennberg JE. [Racial, ethnic, and geographic disparities in rates of knee arthroplasty among Medicare patients](#). *N Engl J Med*. Oct 2 2003;349(14):1350-9.

This paper further illustrates the complexity of the patterns of geographic variation, looking at a single preference-sensitive surgical procedure, arthroplasty (the instillation of an artificial knee) for patients with arthritis of the knee. Striking regional variations are shown on a sex/race-specific basis: in some regions, the rates for black women and non-Hispanic white women are similar, while in others, they are significantly lower; by contrast, among black men, the rates are consistently lower than among non-Hispanic white men. The paper also makes comparisons between Hispanics and non-Hispanics. For both men and women, there is considerable overlap in regional rates.



Evaluating Provider Efficiency in Managing Chronic Illness

Over the years, our research group has paid special attention to the opportunity to develop hospital-specific performance measures for patients with chronic illness. The emphasis has been on measures that compare the efficiency of health care providers in managing patients over time. In the first example, our 1994 Boston-New Haven study, we followed cohorts of patients with hip fractures, strokes, and a few other low-variation conditions forward in time to measure further hospitalizations. We found considerable variation in readmission rates among the hospitals in Boston. (See “[Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven.](#)”)

More recently, we developed a strategy for measuring “relative efficiency” according to the hospital (or health system) used. Instead of following patients forward in time, we used a “follow-back” strategy: all Medicare patients with chronic illness who died were identified and assigned to the hospital they most often used during the last two years of life. Because patients tend to use the same hospital and its associated physicians during this phase in the progression of chronic illness, we could compare them in terms of per decedent spending, resource inputs, utilization, and adherence to quality standards, and thus measure their efficiency in managing severe chronic illness relative to benchmarks provided by “best practice” providers.

The methods and examples of the evaluation strategy were published in three articles in 2004-2005:

- Wennberg JE, Fisher ES, Stukel TA, Skinner JS, Sharp SM, Bronner KK. [Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States.](#) *BMJ*. Mar 13 2004;328(7440):607.

Published in the *British Medical Journal*, this paper compared the use of medical care by patients with chronic illness during the last six months of life among “America’s best hospitals”—77 hospitals that appeared on *U.S. News and World Report’s* 2001 list as the best in delivering geriatric care and in caring for patients with cancer, heart, and lung disease. The conclusion: “Striking variation exists in the utilization of end-of-life care among U.S. medical centres with strong national reputations for clinical care.”

- Wennberg JE, Fisher ES, Stukel TA, Sharp SM. [Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness.](#) *Health Aff (Millwood)*. 2004;Suppl Variation:Var5-18.

Upon further analysis of the variation phenomenon among the chronically ill patients at the nation’s 77 best hospitals, the patterns of care in a hospital were shown to be consistent across type of chronic illness. Hospitals that had high rates for one condition, say cancer, tended to have high rates for other conditions such as congestive heart failure and chronic obstructive pulmonary disease. It was also consistent over time; provider-specific patterns of spending, resource inputs, and utilization seen during the last six months were highly correlated with the patterns in previous six-month periods prior to death. Just as we noted in the Boston-New Haven comparisons, the effect we were measuring was a property of the system: the effect of supply on clinical decision-making.



- Wennberg JE, Fisher ES, Baker L, Sharp SM, Bronner KK. [Evaluating the efficiency of California providers in caring for patients with chronic illnesses](#). *Health Aff (Millwood)*. Jul-Dec 2005;Suppl Web Exclusives:W5-526-43.

This paper applied the follow-back from death cohort design to document longitudinal performance among patient cohorts according to the region where they lived and the specific hospital they most often used. Again, we documented extensive variation among regions and, for the first time, the extensive variation in spending, resource use, utilization, and quality among the patient populations using the various hospitals located within a given region. The importance of a population-based strategy was further emphasized: per capita reimbursement was highly correlated with number of days spent in hospital, much more so than average reimbursement per day spent in hospital. The hospital effect—“the subliminal influence of capacity”—was shown to be much more important than age, race, income, or nature of chronic condition in influencing hospitalization rate.

This paper also put forward the concept of benchmarking as a strategy for evaluating relative efficiency. At the regional level, on the basis of its lower spending, lower resource inputs, and utilization rates, the Sacramento region was selected as regional benchmark. Applying this benchmark to the Los Angeles region identified substantial savings for Medicare that would accrue if Los Angeles were somehow able to achieve the same per capita level of care. At the hospital-specific level, the more than twofold variation in spending and utilization seen among Los Angeles hospitals—all of which exceeded the Sacramento benchmark—pointed to significant opportunities for improving performance.

We defended the strategy of relative efficiency by appealing to the importance of dealing directly with supply-sensitive care. While it would be preferable if this form of care could be based on a detailed understanding of the cost-effectiveness of supply-sensitive care—“for example, whom to hospitalize, when to schedule a revisit or when to refer”—such scientifically valid, detailed evidence defining efficient clinical pathways simply doesn’t exist:

“It will take a long time and a major reorientation of the academic research agenda to provide such clinical evidence, if indeed it is ever possible to do so. In the meantime, we argue that the results of natural experiments—population-based studies comparing the overall quality and outcomes for similarly ill patients exposed to different levels of care intensity—should be used to establish benchmarks of relative efficiency. So far, these studies indicate no marginal gain with greater resource use observed within the United States.”



Recent Research: Risk Adjustment and the Observational Intensity Bias

In preparing the [2008 Dartmouth Atlas](#) on end-of-life care, we noted that adjusting for the number of comorbid diagnoses listed in the claims data resulted in paradoxical changes in estimated spending, resource inputs, and utilization rates, making regions like Miami and Los Angeles appear more efficient than Minneapolis and Seattle. We traced the tendency to list diagnoses in the claims data to the frequency of encounters with physicians, measured as the numbers of physicians visits, referrals, diagnostic tests, and imaging exams. As a temporary fix, we modified our adjustment method to adjust only for the first-listed chronic condition and the presence of one or more comorbidities.

Our concern about the standard methods of risk adjustment increased with the publication of an article by Ong et al[1] that challenged our finding that greater care intensity was not associated with improved outcomes. They followed cohorts of patients with congestive heart failure at six University of California teaching hospitals and Cedars-Sinai Medical Center for up to six months to find that crude mortality rates were similar. However, after adjusting for the number of comorbidities listed in the claims data, significant differences emerged: the hospital with the highest care intensity pattern had mortality rates that were 53% lower than the hospital with the lowest intensity.

To better understand this contradiction in findings, we undertook a series of studies to characterize the association between intensity of patient observation and the frequency of diagnosis and the implications for risk adjustment strategies that depend on the number of comorbidities listed in claims data.

- Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. [Regional variations in diagnostic practices](#). *N Engl J Med*. Jul 1 2010;363(1):45-53.

The first study evaluated a natural experiment that followed Medicare enrollees who migrated from one region to another. We stratified the U.S. into quintiles according to price-adjusted Medicare spending, our measure for the intensity of care. The rates for physician visits, diagnostic testing, and imaging exams were substantially higher among enrollees who migrated from low-intensity regions like Minneapolis or Seattle to high-intensity regions like Miami or Los Angeles than those migrating in the opposite direction. However, baseline illness rates were not associated with migration pattern and mortality following migration was similar, even though, based on count of diagnoses, those moving to higher-intensity regions appeared to become sicker than those moving to less intensive regions.

Our conclusion: Substantial differences in diagnostic practices that are unlikely to be related to patient characteristics are observed across U.S. regions. The use of clinical or claims-based diagnoses in risk adjustment may introduce important biases in comparative effectiveness research, public reporting, and payment reforms.



- Welch HG, Sharp SM, Gottlieb DJ, Skinner JS, Wennberg JE. [Geographic variation in diagnosis frequency and risk of death among Medicare beneficiaries](#). *JAMA*. Mar 16 2011;305(11):1113-8.

This paper further characterized the problem of illness adjustment by showing the effect of diagnosis frequency on case fatality rates. A greater proportion of Medicare enrollees living in regions with high diagnosis frequency are labeled as having a life-threatening chronic illness, as measure by the lezzoni list of conditions. However, the case fatality rates are progressively lower in regions with greater numbers of physician visits, even though population-based mortality rates were the same. Thus, chronic illnesses diagnosed in claims data appear susceptible to the stage migration phenomenon seen in cancer screening: as more people are diagnosed, average severity of illness drops, as do the mortality rates for those with the diagnosis. However, overall population-based mortality remain unchanged.

Our conclusion: Among fee-for-service Medicare beneficiaries, there is an inverse relationship between regional frequency of diagnosis and the case fatality rate for chronic conditions.

- Wennberg JE, Staiger DO, Sharp SM, et al. [Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims](#). *BMJ*. Feb 21 2013;346:f549.

The validity of methods that use diagnoses listed in the claims data to adjust for illness rests on the assumption that recorded diagnoses are independent of supply and thus closely reflect the “true” underlying burden of illness. Our previous studies challenged this assumption and led us to attempt to estimate the bias associated with the frequency of physician visits. In this study, we compared two methods of risk adjustment: the standard method that uses the actual count of diagnoses listed in the claims (such as the Hierarchical Condition Category Score used by Medicare); and a modified method we developed that corrects for the bias associated with observational intensity, as measured by physician visit rate in the last six months of life (a measure that is uncorrelated with age, sex, race-adjusted mortality).

The extent of observational bias became evident by comparing the effect of adjustment on mortality rate. After age, sex, and race adjustment alone, regions with the highest and lowest visit rates had rates that were nearly the same. But after risk adjustment using Medicare’s standard method, regions in the highest quintile for physician visits per capita had mean mortality about 20% lower than regions in lowest quintile—changes from baseline rates that were similar in magnitude to that reported by [Ong et al](#) [2]. However, when risk scores were adjusted to remove observational intensity bias, the adjusted mortality rates became similar in high and low intensity regions.

Our conclusion: The rates of physician visits introduce substantial bias when regional mortality and spending rates are adjusted for illness using comorbidity measures recorded in Medicare claims. Visit corrected comorbidity measures substantially reduce bias.



- Wennberg DE, Sharp SM, Bevan G, Skinner JS, Gottlieb DJ, Wennberg JE. [A population health approach to reducing observational intensity bias in health risk adjustment: cross sectional analysis of insurance claims](#). *BMJ*. Apr 10 2014;348:g2392.

The fourth study in this series compared the claims-based methods of risk adjustment to two new approaches that do not depend on physician diagnoses. The first used a single measure of deprivation: the percent of the population below poverty, as measured by the U.S. Census. The second was a composite index of population health using measures that do not depend on claims data diagnoses: the incidence of hip fractures and strokes measured by hospitalization rate, and obesity, smoking status, and self-reported illness, obtained from an annual population-based survey health conducted by the Centers for Disease Control and Prevention.

Our conclusion: The new approaches, especially the composite index of population health (which accounted for more than 60% of variation in mortality rates), performed substantially better than the claims-based methods in explaining variation in mortality rates and avoided implausible swings in predicted mortality and spending in regions with high and low visit rates.

[2] Ong MK, Mangione CM, Romano PS, Zhou Q, Auerbach AD, Chun A, et al. [Looking forward, looking back: assessing variations in hospital resource use and outcomes for elderly patients with heart failure](#). *Circ Cardiovasc Qual Outcomes*. 2009;2:548-57.



The Policy Papers

Over the years, my colleagues and I have written a number of papers that address the challenge of our studies to conventional assumptions about the nature of supply and demand in the health care economy. Here, more or less in chronologic order, are examples of these papers.

1. [The Need for Assessing the Outcomes of Common Medical Practices](#)
2. [Professional Uncertainty and the Problem of Supplier-induced Demand](#)
3. [Should the Cost of Insurance Reflect the Cost of Use of Care in Local Hospital Markets?](#)
4. [Dealing with Medical Practice Variations: A Proposal for Action](#)
5. [Small Area Analysis and the Medical Care Outcome Problem](#)
6. [What is Outcomes Research?](#)
7. [Are Randomized Controlled Trials Controlled? Patient Preferences and Unblind Trials](#)
8. [Innovation and the Policies of Limits in a Changing Health Care Economy](#)
9. [Avoiding the Unintended Consequences of Growth in Medical Care: How Might More Be Worse?](#)
10. [The Marketplace in Health Care Reform: Demographic Limits of Managed Competition](#)
11. [Finding Equilibrium in the U.S. Physician Supply](#)
12. [Benchmarking the U.S. Physician Workforce](#)
13. [Prioritizing Oregon's Hospital Resources](#)
14. [Regional Professional Foundations](#)
15. [The Randolph Plan](#)
16. [The Hanover Experiment](#)
17. [Geography and the Debate Over Medicare Reform](#)
18. [Toward the Tipping Point: Decision Aids and Informed Patient Choice](#)
19. [Extending the P4P Agenda](#)
20. [An Agenda for Change](#)
21. [Tracking Medicine: A Researcher's Quest to Understand Health Care](#)



The Need for Assessing the Outcomes of Common Medical Practices

- Wennberg JE, Bunker JP, Barnes B. [The need for assessing the outcome of common medical practices](#). *Annu Rev Public Health*. 1980;1:277-95.

In the mid-1970s, I was a member of the Center for the Analysis of Health Practices at the Harvard School of Public Health. It was during this time that John Bunker, Benjamin Barnes, and I undertook a review of the scientific literature to document the clinical controversies concerning the risks and benefits of common surgical procedures. We sought to establish a link between scientific uncertainty and the degree of variation we were observing in surgery rates among small areas. The study uncovered extensive professional disagreement on reasons why particular operations should be performed—even on such issues as whether a given operation was undertaken to increase life expectancy or to improve the quality of life. We also found extensive disagreement as to what the outcomes actually were and their chances of happening. We summarized the implications as follows:

“The limits on informed decision-making in medical markets are more severe than is generally realized ... If the outcomes of alternative treatments are not understood, how is it possible to make informed decisions or give informed consent? ... When decisions are made, whose values are being expressed, the patient’s or the physician’s? The geographic variations in exposure rates are consistent with the thesis that medical care choices are highly dependent on the preferences of physicians. When professional disagreement is strong and patients delegate decision-making to physicians, the probability of exposure will often depend on the style of practice of the physician or clinic selected for care rather than the nature and severity of illness.”

There were important implications for public regulatory policy:

“Much of current [regulatory] policy is based on the belief that an underlying professional consensus exists on the optimum methods for allocating medical technology. [Utilization review programs] assume that meaningful consensus exists on which standards of care [will assure] that only necessary hospitalizations and procedures are undertaken. Similarly, the development of national standards for health resources—physicians per capita and facilities per capita—rests on the assumption that there is a determined relationship between resource input and health care outcomes. However, the existing evidence is usually insufficient to settle controversies on the value of common medical practices. Therefore, consensus standards represent weighted averages of those selected to establish them and cannot serve as a basis for the rational allocation of medical care technology.”

Our study concluded with the recommendation for a substantial public investment in the assessment of the outcomes of medical practices.



Professional Uncertainty and the Problem of Supplier-Induced Demand

- Wennberg JE, Barnes BA, Zubkoff M. [Professional uncertainty and the problem of supplier-induced demand](#). *Soc Sci Med*. 1982;16(7):811-24.

Most policymakers have assumed that in health care markets it is rational for patients to do something that consumers would otherwise never do: delegate the decision on what to buy to the seller. The reasoning goes as follows: patients simply do not know enough to act on their own behalf. They are uncertain about the nature of the disease they may have and the treatment they should use. The special competencies acquired by physicians through their arduous scientific education, their clinical experience, their lifetime commitment to keep up with the progress of biomedical science, and their adherence to professional ethics make it possible for physicians to act as purchasing agents for their patients: to vicariously choose the treatment patients would want, if only they were themselves physicians.

With the growth of systems of entitlements and third-party payments, physicians were asked to play a second crucial role in the health care economy: to act also as an agent for society as the guardian for society's resources, to modulate and control the system's capacity within bounds prescribed by the knowledge of what works and what patients want. While this opens the possibility that the public purposes of health insurance might be subverted by greed, by a supplier-induced demand that prescribes unnecessary care for reasons of professional profit or prestige, professional ethics assures that most physicians will not induce demand; and through patrolling the market with professional peer review, the few outlier physicians who do will be discovered and disciplined.

By the early 1970s, our studies were directly challenging the rational agency theory as a reasonable description of the behavior of the health care economy. Professional uncertainty concerning the prognosis of disease and the value of treatment was much greater than previously assumed; practice variations, unexplained by patient demand, were ubiquitous, better characterized as a central tendency of the market than an outlier phenomenon explained by deviant physicians. In essence, while neo-classical economists had envisioned an economy that through physician agency was brought to equilibrium, the facts seemed better explained by a behavioral model in which physician practice style influences utilization, and practice style is influenced by features of local capacity and/or rationalizing medical theories, often unsupported by high quality evidence.

This article summarized the evidence in support of the professional uncertainty hypothesis and illustrated some of the unintended consequences of public regulatory decisions made under the rational agency assumption. In order to address the problem of supplier-induced demand, public policy for health care should emphasize the need to undertake outcomes research; reform insurance markets to reduce unwarranted transfer payments; and promote the active participation of the patient in the choice of treatment as a specific remedy for unwarranted variations.



Should the Cost of Insurance Reflect the Cost of Use in Local Hospital Markets?

- Wennberg JE. [Should the cost of insurance reflect the cost of use in local hospital markets?](#) *N Engl J Med.* Nov 25 1982;307(22):1374-81.

By the early 1980s, our studies had uncovered striking variations in per capita spending among hospital service areas throughout the six New England states that we couldn't explain on the basis of illness or other factors determining medical need. The differences in per capita spending resulted in cross-community transfer payments in which people living in low-cost, low-use regions ended up paying for a portion of the care received by those living in high-cost regions—a result of the actuarial policies that governed the price of insurance. In this article, I illustrated the problem for those insured by Blue Cross and by Medicare and discussed the implications for equity. I also considered the impact on market behavior when the price of insurance is uncoupled from the rate of consumption. These include the impact on decisions to increase the capacity of the local health care system; distortion of competition; and perpetuation of established patterns of expenditures and transfer payments when efforts to set global budgets or regulate supply ignore variations in spending.

Although written in the early 1980s, the issues discussed in this article remain relevant to the policy discussions of today, including questions concerning the pricing of a voucher under proposals for Medicare reform; the payments to be made to accountable care organizations; and the actuarial policies that establish the price of insurance premiums under the insurance exchanges called for in the Affordable Care Act.



Dealing with Medical Practice Variations: A Proposal for Action

- Wennberg JE. [Dealing with medical practice variations: a proposal for action](#). *Health Aff (Millwood)*. Summer 1984;3(2):6-32.

In 1983, during his tenure as the Ivan F. Boesky Journalist in Residence at Harvard, John Iglehart, the founding editor of *Health Affairs*, became interested in practice variation and decided to dedicate the spring 1984 edition to this topic. He asked me to propose a plan to reduce unwarranted variation. My proposal had three steps:

The First Step: Monitoring Performance in Hospital Markets

I called for wide implementation of routine reports based on small area analysis comparing the per capita performance of local hospital markets. The reports would detail population-based profiles of resource allocation; rates for specific procedures; and benchmark the use of acute care hospitals according to cause of admission. Since these reports can be generated using routine databases, most provided by insurance claims, the system could be put in place rapidly and at relatively low cost.

The Second Step: Dealing with the Effectiveness Problem

Based on experience in Vermont and Maine, I believed that information on small area variation would motivate physicians to review their practices; result in changes in the indications for specific services; and promote a reduction in variation similar to that I had seen for tonsillectomies in Vermont. I also hypothesized that, when the controversies raised by the study of practice variation couldn't be resolved by consensus or critical review of the medical literature, academic medicine would be motivated to undertake the necessary outcomes research. Needed changes in medical education would follow: medical students would be taught the methods of the evaluative sciences so they might assess "the strengths and weaknesses of the various practice styles they will encounter in the course of their training and prepare for their own contributions to resolving clinical uncertainties as practicing physicians."

The Third Step: Dealing with the Cost-Containment Problem

Given the imperative to contain costs, it should be in most peoples' interest to reduce the use of hospitals and stabilize per capita costs by appropriately controlling capacity. But this can only occur with the active involvement of medical leaders. While government officials, managers of benefit plans, and representatives of public and private interest groups can demand change, they lack the detailed understanding on the nature of medical choices required to deal with the necessity of hospitalizations. The third step is thus based on the hypothesis that the necessary professional leadership to reduce overuse will emerge: that given feedback reports that identify hospital markets with costly overuse, hospitalization for high-variation medical conditions will be reduced and capacity contained.

Although the reviews of my plan were quite positive, including one by James H. Sammon of the American Medical Association, thirty years later, the necessary changes have yet to happen. With the exception of the Dartmouth Atlas, routine reports based on claims data have yet to emerge; with a few exceptions, academic medical centers have yet to focus on the science of health care delivery; and we still wait for professional leadership to reduce the overuse of acute care hospitals. It remains to be seen if provisions in the Affordable Care Act will succeed in motivating the necessary change.



Small Area Analysis and the Medical Care Outcome Problem

- Wennberg JE. Small Area Analysis and the Medical Care Outcome Problem. In: [*Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*](#). Agency for Health Care Policy and Research #PB90-101387, Rockville, MD, 1990, p. 177-206.

The paper revisits my “mental model” for interpreting medical variation: a simple model showing how four factors—the incidence of illness, the patient’s decision to access to care, and the physician’s decisions in making a diagnosis and in recommending treatment—contribute to variation. I used the pattern of variation in hospitalization rates for orthopedic injuries to illustrate the power of small area analysis to uncover and characterize the importance of discretionary clinical decision-making in determining the rate of utilization. For patients with hip fracture—a low-variation condition—illness drives utilization; patients uniformly seek care and physicians uniformly make the diagnosis and uniformly hospitalize the patient. By contrast, while patients with fracture of the forearm—a high-variation condition—uniformly seek care, physicians disagree on the need for surgery (and hospitalization). The fracture example has served well in helping physicians tie the facts of variation to the reality of clinical practice as they recognize it from their own experience.



What is Outcomes Research?

- Wennberg JE. [What is Outcomes Research?](#) In: Gelijns AC, ed. *Medical Innovation at the Crossroads, Vol I: Modern Methods of Clinical Investigation*. National Academy Press; 1990:33-46.

By the late 1980s, our outcomes research work in Maine was receiving attention in Washington as a possible strategy for dealing with unwarranted variation in medical practice. As a member of the Institute of Medicine's Committee on Technologic Innovation, I was asked to contribute a chapter on the Maine project, and I took this opportunity to try to define the domain of outcomes research—the strategies for research, the disciplines involved, and the significance of methodology for the emerging understanding of the importance of patient preferences and shared decision-making in making better medical decisions. I also tried to place the evaluative sciences and outcomes research within a broader historical context as a strategy for dealing with supplier-induced demand. I argued that they have emerged as “a response to the intellectual crisis created by the demise of the rational agency theory and the loss of faith that biomedical science, through its own internal logic, assures effective medical practice.” While my appraisal of the prospects for outcomes research in the 1990s turned out to be overly optimistic, the rebirth of medical effectiveness research and the growth in understanding of the importance of shared decision-making under the Affordable Care Act has enhanced the prospects for a new paradigm.



Are Randomized Controlled Trials Controlled? Patient Preferences and Unblind Trials

- McPherson K, Britton AR, Wennberg JE. [Are randomized controlled trials controlled? Patient preferences and unblind trials.](#) *J R Soc Med.* Dec 1997;90(12):652-6.

The importance of active patient choice in determining the medical necessity of preference-sensitive care, such as elective surgery, raised an important question concerning the conduct of outcomes research. Do patient preferences influence the effectiveness of care? If patient preferences can influence the effectiveness of treatments through poorly understood interactions between patient preferences and outcomes, then randomized clinical trials may wrongly attribute treatment effects to physiological rather than psychological mechanisms. Unless we can disentangle these effects, we may never know, even approximately, how much of medicine is attributable to psychological processes. In this paper, Klim McPherson, Annie Britton, and I discuss these implications and examine several proposed clinical trials designed to disentangle preference effects from physiologic effects.



Innovation and the Policies of Limits in a Changing Health Care Economy

- Wennberg JE. [Innovation and the Policies of Limits in a Changing Health Care Economy](#). In: Gelijns AC, ed. *Medical Innovation at the Crossroads, Vol III: Technology and Health Care in an Era of Limits*. National Academy Press; 1992:9-33.

In the debate over health reform in the early 1990s, many feared that cost containment, given the advanced state of medical progress, would inevitably lead to health care rationing. In this essay I challenged this interpretation, updating my argument on the importance of supplier behavior in determining the amount and types of care patients utilize. I argued the case for reducing unwarranted physician influence on the choices patients make by replacing delegated decision-making (the rational agency model) with shared decision-making. In contrast to micro-managed care by insurance companies, whose rules of practice can lead to care that patients don't want, the outcomes research/shared decision-making model promised a strategy for truly rationalizing demand for discretionary surgery and avoiding operating on patients who don't want surgery.

Neither outcomes research nor shared decision-making, however, are sufficient for achieving the overall goals of reducing unwarranted variation in utilization and containing costs. Only about 25% of Medicare spending is for elective surgery. Most of the money goes to manage patients with chronic illness. Most of the care provided to those with chronic illness is not driven by explicit theories of efficacy that can be rationalized by shared decision-making and outcomes research. The frequency of use—the rate of physician visits, hospitalization, admissions to ICUs, imaging exams, and laboratory tests—is under the influence of available resources, but in everyday practice, neither physicians nor patients are aware of capacity's effects on decision-making. To force awareness and contain overall costs, policies that set limits on capacity and budgets are needed.

This paper updates many of the issues raised in our 1982 paper, "[Professional Uncertainty and the Problem of Supplier-induced Demand](#)." The discussion concerning the importance of outcomes research and shared decision-making in establishing "true" demand for discretionary interventions draws heavily on the BPH studies in Maine. The discussion on the need to set direct limits on capacity (and why neither shared decision-making nor micro-management of physician decision-making can do this) draws on studies of variation for medical (non-surgical) conditions, particularly the Boston-New Haven comparisons. The paper sets out several principles to guide debate about strategies to set limits and concludes with a discussion of the implications for innovation.

The challenge of setting limits on physician workforce, facilities, and budgets, while rationalizing the doctor-patient relationship through outcomes research and shared decision-making, became recurrent themes in our policy research and efforts to change clinical practice. The policies of limits, however, need to be crafted to deal with the "sea of uncertainty" that surrounds everyday practice of medicine. Here I invoke the concept of "supply-sensitive care": clinical decisions concerning the frequency of physician visits, imaging exams, laboratory tests, hospitalizations, and intensive care admissions that are best understood as subliminal responses to resource capacity and thus best managed by strategies of limits that focus on resource allocation and budgets.



Avoiding the Unintended Consequences of Growth in Medical Care: How Might More Be Worse?

- Fisher ES, Welch HG. [Avoiding the unintended consequences of growth in medical care: how might more be worse?](#) *JAMA*. Feb 3 1999;281(5):446-53.

This paper focused on the clinical interpretation of the consequences of undisciplined growth in medical care—on the growth of care of uncertain value that may expose patients to risk with no assurance of benefit. In a story targeted at clinicians, Elliott Fisher and Gil Welch drew the distinction between decisions governing the use of discrete interventions, such as a diagnostic tests or surgical procedures, and decisions about whether to increase capacity, such as purchasing an MRI machine, employing another physician, or building more ICU beds. They then analyzed how decisions at either level could lead to harm by overdiagnosis and overtreatment, leading in turn to more worry and disability on the part of patients, more medical errors, and adverse effects.

They then turned to the prospects for reducing the risk of harm from more medical care by addressing four underlying causes.

The first focuses on the failure in clinical logic that ignores the fact that most illnesses exist along a spectrum, and overdiagnosis creates illnesses that are unlikely to affect patient well-being while creating worry and disability. The solution? Study the natural history of increasingly mild diseases as detected by diagnostic technologies and evaluate the benefits and harms of treatment of mild disease.

The second focuses on problems in clinical science that lead to excessive extrapolation from randomized clinical trials. Trials are conducted on a restricted group of patients; can the results be applied to patients encountered in everyday practice? The solution? Answers to skeptical questions: has the benefit been proven for this specific patient? Are the risks of treatment at my center known?

The third focuses on failure to understand the impact of system-level resources on treatment rates and outcomes. The solution? Consider the impact of change in capacity when planning to build beds, hire physicians, invest in imaging exams, etc.

The fourth focuses on the tendency to look for more care to be better (and to discount evidence for harm). The solution? Promote the null hypothesis: that interventions should be proven effective, rather than assumed to be effective. Consider seriously the possibility that harm may occur from more medical care.

Much of the research of our group can be viewed as an effort to reduce the unintended consequences of health care outlined by Fisher and Welch: evaluating the “more is better” hypothesis; establishing “which rate is right” for surgery through outcomes research and shared decision-making; evaluating the impact of system capacity on health outcomes; and providing performance measures to help policymakers and system managers promote efficiency and reduce the potential for harm from overuse.



The Marketplace in Health Care Reform: The Demographic Limitations of Managed Competition

- Kronick R, Goodman DC, Wennberg J, Wagner E. [The marketplace in health care reform. The demographic limitations of managed competition.](#) *N Engl J Med.* Jan 14 1993;328(2):148-52.

In the early 1990s, health care reform was once again at the top of the national agenda, and the odds-on favorite to gain approval by the U.S. Congress was the competing HMO model espoused by the Jackson Hole Group. As a sometime member, I became familiar with the “managed competition” model and was particularly drawn to the idea of competition between large capitated group practices such as Kaiser Permanente and Group Health. Such organizations, which I labeled Classic HMOs or C-HMOs, serve a defined population and thus can practice “private sector” health planning with regard to per capita allocation of resources—for example, the number of physicians they hire and the hospital resources they use. They thus “solve” the problem of setting limits on capacity that we viewed as an essential step in cost containment. However, for managed competition to exist (and the need for regulation reduced), there must be more than one competing organization within a given medical market. The most effective competition occurs when all the doctors within a community are grouped into several group practices.

This paper, published in the heat of the national debate over the Clinton health care plan, examined the demography of health care markets to estimate what proportion of the U.S. population lived in areas that might successfully support managed competition. Large segments of the U.S. population live in regions where competition between population-based, fully integrated C-HMOs isn’t feasible. We estimated that only 42% lived in market areas that would support three or more full service C-HMOs; the rest lived in markets where specialty services would need to be shared if the C-HMO benchmarks for efficient practice were to be obtained.

The arguments in the paper remain relevant to today’s debate over the roles of competition and regulation in controlling health care spending.

Our estimates are a further example of the use of population-based measures of the physician workforce to benchmark resource requirements for efficient practice. In subsequent policy research, the resource allocation benchmarks from organized group practice (see *Tracking Medicine*, pages 184-187) play an important role in estimating how many physicians, hospital and ICU beds the nation would need and how much we would need to spend if care were based on integrated delivery systems. And the limitation imposed by the population size on managed competition as the driver for development of integrated, population-based care would be addressed by seeking an alternative model: what we came to call the [Randolph Project](#).



Finding Equilibrium in the U.S. Physician Supply

- Wennberg JE, Goodman DC, Nease RF, Keller RB. [Finding equilibrium in U.S. physician supply.](#) *Health Aff (Millwood)*. Summer 1993;12(2):89-103.

This paper focused on physician workforce policy in light of the likely passage of the Clinton health care plan with its provision for managed competition among HMOs and fee-for-service practices. It begins by examining the workforce requirements of the classic staff model HMO or “C-HMO.” As Alain Enthoven noted, in determining the resources required, C-HMOs practice “private sector health planning”: they plan their workforce in terms of the size of the population they serve. This leads to a strong competitive advantage over fee-for-service organizations: its workforce and hospital bed to population ratios are set lower than those that exist among its fee-for-service competitors. If managed competition were to succeed in reducing overcapacity toward the HMO benchmark, American physicians would face massive unemployment. However, for reasons delineated in the paper, we did not believe that managed care was a reliable way for “clearing the market” of excess capacity.

Instead, we argued that for the “market” to work, we needed to complement the C-HMOs’ commitment to private health planning with public sector health planning to deal with excess capacity in a way which minimizes supplier-induced demand, fosters improvement in health care delivery through outcomes research, and promotes innovation, while at the same time brings the supply of clinically active physicians more into balance with the numbers of physicians required by C-HMOs. We presented a national workforce plan that specifies the possible points of intervention to obtain balance and outlines the types of interventions that might be considered.

In the final section of this paper, we approached the question of how to implement our plan. In parts of the country where population size can support competition between classic HMOs, private sector health planning under the rules that would govern managed competition might work best. See “[The Marketplace in Health Care Reform: The Demographic Limits of Managed Competition.](#)” However, elsewhere, where the population base is too small to support such competition, new public policy thinking is required. We suggested a regional organization be formed to take responsibility for managing the workforce plan. See “[Regional Professional Foundations.](#)”



Benchmarking the U.S. Physician Workforce

- Goodman DC, Fisher ES, Bubolz TA, Mohr JE, Poage JF, Wennberg JE. [Benchmarking the U.S. physician workforce. An alternative to needs-based or demand-based planning.](#) *JAMA*. Dec 11 1996;276(22):1811-7.

Given the nature of supply and demand in medical markets, decisions governing the size and specialty composition of the workforce relative to the population served are key in determining the type and quantity of care Americans consume as well as the amount of money spent on health care. To an important extent, the per capita numbers and the specialty composition of the U.S. physician workforce are determined by policies of federal and state governments through decisions affecting post-graduate training of physicians, support to medical schools, and immigration policies. But how many physicians are needed? Traditionally, the need for physicians has depended on either needs-based or demand-based planning. In this article, we argue that weaknesses in the scientific basis for clinical decision-making and the pervasive influence of supply in inducing demand for health care represent a serious challenge to these methods of determining need. As an alternative, we suggested in this paper that per capita number of physicians required for the nation (and the mix among specialties) be based on the needs of organized health care systems. If the goal of public policy is to promote the growth of organized systems of care, particularly capitated group practice, then benchmarks based on the per capita requirements of such organizations seemed the best way to determine “a reasonably sized per capita workforce.”

While estimating workforce need is important for state and national planning purposes, this paper also emphasized the importance of local decision-making with regard to hiring practices. Governing the size of the physician workforce relative to size of the population served is key to managing patient care and costs. The use of population-based data in making decisions on hiring physicians first emerged during my work in Vermont. Estimates of physicians labor inputs are now part of the routine reports provided under the Atlas project. Accountable care organizations, because they serve defined populations, should find population-based benchmarking useful in targeting their employment practices toward “a reasonably sized per capita workforce.”



Prioritizing Oregon's Hospital Resources

- Fisher ES, Welch HG, Wennberg JE. [Prioritizing Oregon's hospital resources. An example based on variations in discretionary medical utilization](#). *JAMA*. Apr 8 1992;267(14):1925-31.

In the early 1990s, the State of Oregon was locked in a bitter debate over their plan to expand Medicaid coverage to the uninsured. To determine which services would be covered, a committee of experts was asked to establish priorities for over 700 “condition-treatment pairs” such as bypass surgery for left main disease of the coronary artery of the heart or tonsillectomy for tonsillitis, ranking them from highest to lowest priority. Based on the estimated cost of their delivery, the Oregon state legislature voted to ration care: to fund only those services that ranked 587th or higher.

Concerned over the ethics of rationing treatments from which some would benefit, the paucity of outcomes-based information on which to rank the value of specific treatment-condition dyads, and unnecessary regulatory intrusion into clinical practice, we proposed an alternative plan for reallocating resources to extend entitlement. Based on our studies of the epidemiology of hospitalizations, we proposed that the resources come from reducing the overuse of hospitals for high variation medical conditions. (See “[Will Payment Based on Diagnosis-Related Groups Control Hospital Costs?](#)”)

This approach, while recognizing the need to set limits and priorities, focused on units of supply—not expert ranking of the relative value of specific treatments, where demand should be based on patient preference. The approach is an example of benchmarking in which the task is to target a reduction in capacity sufficient to produce the savings required to meet the goal of policy makers—with capacity reduction targeted to specific hospitals that overuse supply-sensitive care compared to a regional benchmark. The paper illustrates the application of this method to Oregon hospitals and discusses why we believe it achieves the purposes of reallocation without explicit rationing of specific treatments that may benefit patients. Setting limits on units of supply is a strategy we continue to favor in managing the overuse of supply-sensitive care. (See pages 5-8 in “[An Agenda for Change. Improving Quality and Curbing Health Care Spending: Opportunities for the Congress and the Obama Administration](#)”.)



Regional Professional Foundations

- Wennberg JE, Keller R. [Regional professional foundations](#). *Health Aff (Millwood)*. Spring 1994;13(1):257-63.

In the early 1990s, as Clinton’s Health Security Act took final shape, we worked with Senator George Mitchell, then Senate Majority Leader, to include Section 5008 in the legislation, a provision that supported physicians to form Regional Professional Foundations (RPFs). Patterned in part after the Maine Medical Assessment Foundation, RPFs were designed to support the profession in the tasks of health reform in regions where the population was too small to support competition among integrated systems of care. (See “[The Marketplace in Health Care Reform: The Demographic Limits of Managed Competition](#).”)

Bob Keller and I described the RPF concept in a 1994 article whose publication corresponded roughly with the date when Congress finally rejected the Clinton plan. We argued that professional activities to develop shared decision-making, improve quality, advance the scientific basis for clinical decision-making, and support lifetime professional learning should be viewed as intrinsic to rational management of modern health care systems. In less populated regions, they should be supported by public policy. As illustrated by the relationship between PORTs and the Maine Medical Assessment Foundation, RPFs would serve as the essential link between academic researchers and clinical practice in conducting outcomes research. They can also serve as focus for contracting or otherwise funding physicians to undertake new tasks not traditionally paid for under fee-for-service, including work in quality improvement, outcomes research, and for providing services that are more efficiently delivered at the community rather than the individual practice level, such as immunizations or health education programs.

Although the RPF idea died with the Clinton health plan, the problem we sought to address—the lack of infrastructure to support reform of fee-for-service medicine—persists. Another failed attempt emerged later in the 1990s as part of the Randolph Plan. Accountable care organizations, if they are to succeed, will surely require professional infrastructures to support informed choice, practice-based learning and strategies for mitigating supplier-induced demand.



The Randolph Plan

The movement toward managed care did not end with the defeat of the Clinton health plan. Throughout the 1990s, private insurance companies sought to contain costs by forming “preferred provider organizations” (PPOs). The PPO business model focused on fee-for-service networks of physicians with whom the insurance companies negotiated prices and controlled volume by imposing practice guidelines and requiring pre-admission review before paying for elective surgery. PPOs were formed without regard to natural medical communities: physicians typically belonged to multiple networks, each with its own rules for controlling costs; and patients were often forced to change physicians when their doctor wasn’t a member of their insurance company’s PPO. Unlike the Clinton plan, the PPO model provided no incentives for providers to integrate care and manage capacity using private sector health planning. By the end of the decade, its negative effects on patients as well as physicians brought the entire HMO movement into disrepute.

The Randolph Plan, developed in conjunction with the administration and professional staff of Gifford Memorial Hospital in Randolph, Vermont, was proposed as an alternative model for implementing managed care that respected the demography and the historic patterns of practice within local and regional market. It was based on a method for identifying medical markets where cooperative rather than competitive strategies for implementing managed care would likely be required, and designed specifically to provide remedy for unwarranted variation and undisciplined growth in health care delivery by achieving the following steps:

- *An informed patient:* Provide information and counseling to patients concerning access to care, treatment options, and importance of patient preference in deciding among treatment options.
- *Shared decision-making:* When patient preferences for specific treatments vary, structure the doctor-patient relationship to involve the patient directly in the decision process.
- *Best-practices benchmark for capacity:* Examine existing patterns of resource allocation to identify excess capacity (for example, hospital service areas with similar populations, lower beds per 1,000 and equal or better outcomes).
- *Prospective budgeting:* Introduce a system-wide prospective budget strategy providing needed flexibility to reallocate resources to new tasks not covered in the fee structure and to optimize financial incentives for introducing shared decision-making.
- *Reallocate excess capacity to improve population health:* Reallocate resources to services that work and patients want.
- *Integrate care:* Integrate local and regional health care systems by bringing primary and referral care under panel and disease management protocols.
- *Improve the scientific basis of clinical decision-making:* Evaluate the outcomes of care by following up patients according to services used.

The Randolph Plan as proposed to the Health Care Finance Administration (HCFA, now CMS) had three components:



The Financial Plan to establish a budget approach to a defined patient population based on a “virtual capitation” form of prospective reimbursement for all services provided by the local and regional provider group. The financial plan, by moving away from fee-for-service medicine, established incentives to (1) constrain global (total per capita) costs in the area and (2) promote the reallocation of resources to more efficient purposes.

The Patient Care Management Plan to move beyond the de facto pattern of local and regional care to an integrated strategy that links preventive, primary, and specialty care; it developed the benchmarking strategies for guiding resource allocation and established the clinical protocols that guide the implementation of regional disease management approaches, including shared decision-making.

The Population and Health Systems Monitoring System to support (1) the patient care management plan, providing benchmarks and detailed information on resource allocation, utilization and outcomes; and (2) the financial plan, providing information on per capita reimbursements, stability of patient loyalty, market share, and out-of-plan use of care. The monitoring system, based on a modification of the Medicare claims data, would provide information to HCFA, as well as serve as a source of feedback to participants in the pilot project.

Although the Randolph Plan was never implemented, its concepts and ideas motivated Senator Jeffords to introduce legislation supporting demonstration projects that became law as Section 646 of the Omnibus Bill of 2003. (For a brief history of Section 646, see the 2004 Perspective in *Health Affairs* by Paul Harrington, who served as chief of staff for Senator Jeffords.) The strategies outlined in the plan for virtual capitation, shared savings, reallocation of resources, and shared decision-making play a prominent role in the Affordable Care Act: in particular, its accountable care organization (ACO) provision.



The Hanover Experiment

A distinguishing feature of the American culture is its willingness to experiment to figure out what works, even when this means a departure from tradition. Over the past twenty years, an example of radical redesign has taken place in a continuing care retirement community in Hanover, New Hampshire. A collaboration between community members and primary care providers resulted in a redesign of medical practice around the need for care across the life span that resulted in lower hospitalization rates and fewer physician visits compared to controls, while meeting the goals of patients.

Julie Bynum and colleague reported on the natural experiment in *Health Affairs*:

- Bynum JP, Andrews A, Sharp S, McCollough D, Wennberg JE. [Fewer hospitalizations result when primary care is highly integrated into a continuing care retirement community](#). *Health Aff (Millwood)*. May 2011;30(5):975-84.

Further details are reported in *Tracking Medicine* (see pages 221-223).



Geography and the Debate Over Medicare Reform

- Wennberg JE, Fisher ES, Skinner JS. [Geography and the debate over Medicare reform](#). *Health Aff (Millwood)*. Jul-Dec 2002;Suppl Web Exclusives:W96-114.

This 2002 paper introduced the policy world to the “categories of care” concept that had emerged from our studies of the epidemiology of medical care. Medical care can be segregated into three more or less distinct sets of services: effective care, preference-sensitive care, and supply-sensitive care. This distinction is important because the causes and remedies of unwarranted variation differ according to category.

Effective care refers to services of proven effectiveness that involve no significant tradeoffs—all patients with specific medical needs should receive them. Conflict between patients and providers is minimal. The policy problem is underuse and the evidence that greater spending doesn’t lead to better quality.

Preference-sensitive care involves tradeoffs; decisions should therefore be based on patient preferences and values. Elective surgeries are typical examples. Although opinions are strongly held by clinical advocates, supporting scientific evidence may be weak or strong. The effect of supply of resources is variable. Patient and provider values are often in conflict. The policy problem is misuse: unwarranted variation in elective treatment that reflects provider rather than patient preference for care. Reducing unwarranted variation (learning which rate is right) requires the active involvement of the patient in choice of treatment for preference-sensitive care.

Supply-sensitive care is generally provided in the absence of specific theories of benefit that govern the frequency of use. What is at stake is the intensity of everyday care, typically in the treatment of medical (non-surgical) conditions. Clinical science provides little or no guidance on when to schedule a revisit, perform a diagnostic test, hospitalize, or admit to intensive care. However, utilization is strongly influenced by supply of resources; in some cases, patient preferences and values should play a central role, particularly for end-of-life care. The policy problem is primarily overuse, particularly at the end of life. Reducing unwarranted variation in supply-sensitive care is doubly important because variation in this category of care is the principal reason for the more than twofold variation in Medicare spending among regions. Reducing variation requires that health care organizations become accountable for capacity and budgets.

The paper recommended that Medicare undertake a national program to establish “Comprehensive Centers for Medical Excellence” to address underlying causes of variation in each of the three categories of care. Congress gave CMS the legislative authority to undertake such a demonstration in Section 646 of the Omnibus Bill of 2003.



Toward the Tipping Point: Decision Aids and Informed Patient Choice

- O'Connor AM, Wennberg JE, Legare F, et al. [Toward the 'tipping point': decision aids and informed patient choice](#). *Health Aff (Millwood)*. May-Jun 2007;26(3):716-25.

This 2007 paper summarized progress in implementing informed patient choice as the standard of practice for preference-sensitive conditions. Beginning with the emphasis on shared decision-making by the October 1982 President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research and the early research by the PORT teams that established the central importance of patient preferences in determining the medical necessity of common surgical interventions, research and development projects over the past thirty years on many fronts have succeeded in building an infrastructure to support shared decision-making, including decision aids and instruments to measure patient decision quality. However, to reach the “tipping point”—the place where informed patient choice would emerge as a standard of practice—significant cultural, regulatory, legal, and economic barriers must be overcome.

We argue that the policy community should develop a new standard for defining medical necessity. The current standard approach used by insurance companies and Medicare is based on the physician as the determinant factor. Under our suggestion, the payer would continue to determine what is covered and the profession (and the research community) would determine which treatment options are reasonable for preference-sensitive conditions. But the informed patient would determine medical necessity. Economic incentives should be developed to support shared decision-making and encourage physicians as well as patients to engage in it. Finally, we argue that informed patient choice should become the legal standard for meeting requirements for informed consent.

Health plans and employers need to develop new payment strategies that encourage both patients and providers to participate in shared decision-making. The legal profession needs to recognize the limitations of traditional informed consent in assuring shared decision-making and adopt a new standard based on the requirements for achieving informed patients through shared decision-making.



Extending the P4P Agenda

In the run-up to the 2008 presidential election, the nation once again seemed to be preparing for reform of the U.S. health care system. A prominent theme was pay-for-performance (P4P)—change the reimbursement system to reward physicians for value, not just utilization. A series of “proof of concept” pilot projects were underway. But they were narrowly focused on underuse—on ways to increase the use of services in the effective care category. While reducing underuse is incredibly important—it could save lives—only a small percentage of Medicare services and spending are for effective care. To address the ethical problems of misuse of elective surgery and other preference-sensitive care—to reduce the medical error of providing care to the wrong patient—P4P needed to be extended to include this category of care. To address the problem of overuse of physician and hospital care in managing acute and chronic illness—and the associated twofold variation in Medicare spending among regions and hospitals—P4P also needed to be extended to include supply-sensitive care.

In the August 2007 issue of *Health Affairs*, we published a two-part proposal on ways Medicare could extend P4P to these categories of care.

- Wennberg JE, O'Connor AM, Collins ED, Weinstein JN. [Extending the P4P agenda, part 1: how Medicare can improve patient decision making and reduce unnecessary care](#). *Health Aff (Millwood)*. Nov-Dec 2007;26(6):1564-74.

This paper outlined a strategy to achieve informed patient choice as the standard of practice with Medicare taking the lead.

- Wennberg JE, Fisher ES, Skinner JS, Bronner KK. [Extending the P4P agenda, part 2: how Medicare can reduce waste and improve the care of the chronically ill](#). *Health Aff (Millwood)*. Nov-Dec 2007;26(6):1575-85.

This paper outlined a strategy for improving the scientific basis for managing chronic illness and for Medicare to extend its pay-for-performance agenda to ensure that, within ten years, all Americans with severe chronic illness have access to accountable health care organizations.



An Agenda for Change: Improving Quality and Curbing Health Care Spending. Opportunities for Congress and the Obama Administration

- Wennberg JE, Brownlee S, Fisher ES, Skinner J, Weinstein JN. [An Agenda for Change: Improving Quality and Curbing Health Care Spending. Opportunities for Congress and the Obama Administration.](#) Lebanon, NH: 2008; The Dartmouth Institute for Health Policy and Clinical Practice and The Trustees of Dartmouth College.

With the election of President Obama in November 2007, the nation was once again on a track toward health care reform. In this 2008 Dartmouth Atlas white paper, we summarized major findings and policy implications from our research into the causes and remedies of unwarranted variations in health care delivery. It identified the following key shortcomings in our health care system:

- Clinical decisions that fail to adequately take patient preferences into account, resulting in unnecessary, unwanted elective surgery;
- Poor clinical science;
- Workforce policies that result in undersupply of primary care physicians and oversupply of specialists;
- Insurance markets ill equipped to address unwarranted variation because prices of premiums are not closely linked to local cost of medical care;
- Disorganized, poorly coordinated, and inefficient care resulting in overuse of supply-sensitive care: physician visits, referrals, hospitalizations, and stays in ICUs, particularly in treatment of chronic illness.

The intensity of the frequency of contact with physicians affects the number of diagnoses patients get, and thus how sick they are thought to be. In response, Medicare pays more to high-use regions because their patients appear to be sicker; increases transfer payments from more efficient to high-use regions; and induces further increases in utilization and apparent sickness in high-use regions by providing more money to hire more physicians and build more beds, thus increasing still further the intensity of care, and thus the number of diagnoses and apparent sickness, and so on in a vicious circle, with no apparent benefit in terms of health.

The Affordable Care Act, passed and signed into law in March 2010, includes several provisions that, if successfully implemented, should help remedy these shortcomings. The primary care “medical home,” the accountable care organization, and bundled payments are designed to promote organized care. The Patient-Centered Outcomes Research Institute is charged with responsibility for improving the scientific basis for clinical decision-making. While various provisions promote shared decision-making, they are still unfunded; however, the CMS Innovation Center is investing resources in shared decision-making demonstration projects. A Workforce Commission has been established that could become a focus for challenging current workforce policy. The Affordable Care Act also established regulatory authority through the Medicare Advisory Board to set limits on spending, but whether this controversial provision will be implemented remains to be seen.



Editorials, Commentaries, & Perspectives

Over the years, I have written or co-authored a number of editorials, often in the form of an invited commentary on an article that appeared in the same issue of the medical journal. Here is a selection of these editorials, listed chronologically by date of publication.

1. Bunker JP. Editorial: [Operation rates, mortality statistics and the quality of life](#). *N Engl J Med*. Dec 6 1973;289(23):1249-51.
2. Wennberg JE. [Physician uncertainty, specialty ideology, and a second opinion prior to tonsillectomy](#). *Pediatrics*. Jun 1977;59(6):952.
3. Wennberg JE. Wennberg: [Administrators must pursue policies that reduce excess beds, employees](#). *Health Manage Q*. 1984-1985 Winter:6-7.
4. Wennberg JE. [Practice variations: why all the fuss?](#) *Internist*. 1985 Apr;26(4):6-8.
5. Wennberg JE. [On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices](#). *Med Care*. May 1985;23(5):512-20.
6. Wennberg J. [Which rate is right?](#) *N Engl J Med*. Jan 30 1986;314(5):310-1.
7. Wennberg JE. [Setting outcome-based standards for carotid endarterectomy](#). *JAMA*. Nov 14 1986;256(18):2566-7.
8. Wennberg JE. [Population illness rates do not explain population hospitalization rates. A comment on Mark Blumberg's thesis that morbidity adjusters are needed to interpret small area variations](#). *Med Care*. Apr 1987;25(4):354-9.
9. Wennberg JE. [When should patients be hospitalized?](#) *Internist*. May-Jun 1987;28(5):32.
10. Wennberg JE. [The paradox of appropriate care](#). *JAMA*. Nov 13 1987;258(18):2568-9.
11. Wennberg JE. [Improving the medical decision-making process](#). *Health Aff (Millwood)*. Spring 1988;7(1):99-106.
12. Wennberg JE. [Non-experimental methods in the assessment of efficacy](#). *Med Decis Making*. Jul-Sep 1988;8(3):175-6.
13. Wennberg JE. [Outcomes research: the art of making the right decision](#). *Internist*. Jul-Aug 1990;31(7):26, 28.
14. Wennberg JE. [Outcomes research, cost containment, and the fear of health care rationing](#). *N Engl J Med*. Oct 25 1990;323(17):1202-4.
15. Wennberg JE. [Unwanted variations in the rules of practice](#). *JAMA*. Mar 13 1991;265(10):1306-7.



16. Wennberg JE. [Outcomes research, patient preference, and the primary care physician](#). *J Am Board Fam Pract*. Sep-Oct 1991;4(5):365-7.
17. Wennberg JE. [The wrong way to heal Medicare](#). *The New York Times*. October 6 1995;Section A:31.
18. Wennberg JE. [On the appropriateness of small-area analysis for cost containment](#). *Health Aff (Millwood)*. Winter 1996;15(4):164-7.
19. Wennberg JE. [Medical Necessity and the Debate over \[Expletive Deleted\] Care](#). *Public Health Rep*. Jul 1997;112(4):306-307.
20. Lundberg GD, Wennberg JE. [A JAMA theme issue on quality of care. A new proposal and a call to action](#). *JAMA*. Nov 19 1997;278(19):1615-6.
21. Sackett DL, Wennberg JE. [Choosing the best research design for each question](#). *BMJ*. Dec 20-27 1997;315(7123):1636.
22. Weinstein JN, Goodman D, Wennberg JE. [The orthopaedic workforce: which rate is right?](#) *J Bone Joint Surg Am*. Mar 1998;80(3):327-30.
23. Wennberg JE. [Understanding geographic variations in health care delivery](#). *N Engl J Med*. Jan 7 1999;340(1):52-3.
24. Goodman DC, Wennberg JE. [Maps and health: the challenges of interpretation](#). *J Public Health Manag Pract*. Jul 1999;5(4):xiii-xvii.
25. Wennberg JE, Fisher ES. [Finding high quality, efficient providers for value purchasing: cohort methods better than methods based on events](#). *Med Care*. Oct 2002;40(10):853-5.
26. O'Connor AM, Mulley AG, Jr., Wennberg JE. [Standard consultations are not enough to ensure decision quality regarding preference-sensitive options](#). *J Natl Cancer Inst*. Apr 16 2003;95(8):570-1.
27. Skinner J, Wennberg JE. Exceptionalism or extravagance? [What's different about health care in South Florida](#). *Health Aff (Millwood)*. Jul-Dec 2003;Suppl Web Exclusives:W3-372-5.
28. Wennberg DE, Wennberg JE. [Addressing variations: is there hope for the future?](#) *Health Aff (Millwood)*. Jul-Dec 2003;Suppl Web Exclusives:W3-614-7.
29. Wennberg JE. [Practice variations and health care reform: connecting the dots](#). *Health Aff (Millwood)*. 2004;Suppl Variation:Var140-4.
30. Wennberg JE. [Time to tackle unwarranted variations in practice](#). *BMJ*. 2011 Mar 17;342:d1513.
31. Wennberg JE. [Forty years of unwarranted variation--and still counting](#). *Health Policy*. Jan 2014;114(1):1-2.