

There's No Reason for It, It's Just Our Policy

Il n'y pas vraiment de raison,
c'est simplement notre politique

by ROBERT G. EVANS

Abstract

On June 1, 2009 the town of McAllen, Texas, rose to brief prominence on the American political stage. With the highest (bar Miami) per-beneficiary costs in the entire US Medicare program, it was featured in an essay in *The New Yorker* by Atul Gawande, then seized upon by President Obama: "This is what we have to fix." Behind the headlines were decades of documentation of clinical practice and analysis of regional variations by John Wennberg, Elliott Fisher and their colleagues, and by Leslie and Noralou Roos and theirs. The implications for health systems were grasped over 30 years ago and have been confirmed by more recent work. Efforts to understand these variations within standard economic theory have, however, had limited success.

Résumé

Le 1er juin 2009, la ville de McAllen, Texas, a fait la manchette sur la scène politique aux États-Unis. Elle présentait les plus hauts coûts par bénéficiaire (mise à part Miami) du régime d'assurance-maladie (Medicare) aux États-Unis. Le cas de McAllen a fait l'objet d'un article écrit par Atul Gawande dans le *New Yorker*, puis a été repris par le président Obama comme exemple de ce qui « doit être corrigé ». Derrière ces grands titres, il y avait des années de documentation sur la pratique clinique et d'analyses sur les variations régionales effectuées par John Wennberg, Elliott Fisher et leurs collègues et par Leslie et Noralou Roos et leurs collègues.

Les implications pour le système de santé ont été dégagées il y a plus de 30 ans, puis confirmées par des travaux plus récents. Cependant, les tentatives pour comprendre ces variations dans le cadre des théories de l'économie ont connues bien peu de succès.

WHEN MY DAUGHTER WAS BORN IN 1966, MOTHER AND BABY SPENT FIVE days at the Boston Lying-in Hospital – standard for an uncomplicated delivery. Had I gone, in 1964, to Berkeley instead of Harvard, they would have stayed three days. This east–west differential was well known, and there was no evidence of poorer outcomes in the Bay area. The potential for savings in bed-days and money in the Boston hospital system were obvious – normal deliveries were the largest single category of admissions. But no one in authority seems to have taken any interest. Their priorities were elsewhere: scrambling for the serious federal money beginning to flow from the new Medicare and Medicaid programs.

Forty years on, geographic variations in health services use have a somewhat higher profile in the United States. Atul Gawande (2009), writing in *The New Yorker*, has just provided an example of “knowledge transfer” beyond the wildest dreams of other health services researchers. His essay on the remarkable state of health services in McAllen, Texas, an otherwise ordinary town on the Mexican border, was immediately seized upon by President Obama and put before his staff and leading congressional Democrats: “This is what we’ve got to fix” (Pear 2009).

McAllen has the second-highest per capita Medicare expenditures of any region in the United States, nearly double the national average and double those in the very similar town of El Paso, farther along the border.¹ It provides an arresting snapshot, from the broader picture, of very large regional variations in use and costs that are unrelated either to patient needs or to health outcomes. Gawande’s conversations with local doctors turned up the usual suspects – sicker patients, better-quality care, threats of malpractice litigation; none held water. Finally a surgeon, with refreshing candour, cut in: “Come on... we all know these arguments are bullshit. There’s overutilization here, pure and simple.”

His interpretation would no doubt be contested by representatives of the local medical community. What is not contestable is the simple fact. Patterns of medical practice, reflected in per capita rates of service use and expenditure, vary widely across different regions, and no satisfactory explanations, in terms of patient needs or health outcomes, have ever been offered. The routine responses by apologists for the status quo are variants on those pungently characterized by the Texas surgeon.

These regional variations have been patiently tracked through a generation of research by John Wennberg and his colleagues at the Dartmouth Medical School. Their increasingly comprehensive data collection, sophisticated analysis and effective

communication have built up an ever more compelling case that such variations reflect inappropriate servicing – simple wasted effort – on a very large scale. That case is increasingly being heard: “the research by Dartmouth experts who have documented wide geographic variations in health spending ... has become phenomenally influential on Capitol Hill ...” (Pear 2009).

Peter Orszag, President Obama’s budget director and former director of the Congressional Budget Office, has repeatedly pointed out that the greatest threat to the fiscal stability of the United States is posed by rising health services costs (Orszag 2008; Orszag and Ellis 2007). He has highlighted the central fact of very large regional variations in per-enrollee costs. The Gawande essay was not a complete surprise to the president.

Wennberg’s professional colleagues have also recognized the significance of the Dartmouth program. In 2007, the leading American health policy journal *Health Affairs* named him the most influential health policy researcher of the past 25 years. In 2008, the Institute of Medicine presented Wennberg with the Gustav O. Lienhard Award “for his leading role in reshaping the US health care system to focus on objective evidence and outcomes rather than physician preference as the basis for treatment decisions ...” (Institute of Medicine 2009).

The honours are unquestionably richly deserved. No one could deny the massive impact of the Dartmouth studies on how health services researchers – and increasingly, policy makers – understand the determinants and effects of medical care, not just in the United States but over much of the high-income world. But the Lienhard Award citation is, unfortunately, premature.

McAllen reminds us that Wennberg’s impact on medical practice and patient care is much harder to find. The American political response to President Obama’s championing of Gawande has been profoundly perverse. Representatives of high-spending states such as Massachusetts and New York have dismissed the Dartmouth data as inconclusive; representatives of low-spending states have welcomed the demonstration that they were being short-changed by Washington and deserved more federal money (Pear 2009).

The usual apologists for American health services have taken up the usual pre-prepared positions and begun a powerful campaign to discredit or at least to confuse and distract from the evidence, and in any case to frustrate any effort to build a rational policy response. Rather than an “outstanding achievement in improving health care services in the United States,” the variations research may well sink from sight as Washington moves on to the next burning issue.

This much is the daily news. There are, however, three themes that may not be immediately obvious from the current discussion. First, the principal messages from Gawande’s powerful essay have been available for at least 30 years. They have had no impact on health policies for the same reasons that they are likely to be dismissed

now. Second, the efforts by economists to understand clinical variations within the framework of standard or “mainstream” economic theory have been as jejune and as unsuccessful as those by spokesmen for the medical community. And finally, large geographic variations in clinical practice are not a peculiar consequence of the bizarre American financing system. They are found everywhere. In particular, they are found in Canada, where they could provide a powerful counterpoint to the endless claims of “underfunding” and “shortage” – if anyone in authority were paying attention.

A remarkable early finding was the “surgical signature” (Wennberg and Gittelsohn 1973, 1982). Comparisons of surgical rates among small areas showed that they were not uniformly high or low. A region might have a relatively high rate on one procedure, but be low on another.

... large geographic variations in clinical practice are not a peculiar consequence of the bizarre American financing system. They are found everywhere.

Furthermore, these patterns were associated with particular surgeons; if a surgeon moved from one region to another, the pattern of rates moved with him. Clinicians have different perceptions

as to the relative value or effectiveness of particular procedures, independently of the underlying evidence, which may be masked in aggregate comparisons.

Similar findings emerged from the Manitoba research group led by Leslie and Noralou Roos. Their studies of tonsillectomy identified “believers” and “non-believers” among physicians, as reflected in their rates of performance of the procedure or referral for it (Roos et al. 1977). Other Manitoba studies identified “hospital-prone” physicians, who were on average much more likely to admit patients for a given problem and set of patient characteristics (Roos et al. 1986).

Moreover, when a new surgeon moved into an area, the workloads of established surgeons did not fall. Rather, total surgical rates rose to accommodate the new capacity. But when a surgeon left, the workloads of the remaining surgeons rose to maintain the established population rate. The authors’ best explanation for observed population surgical rates was simply physician discretion (Roos 1983).²

To return to the Dartmouth data, the BPH (benign hyperplasia of the prostate) studies traced variations in surgical rates to surgeons’ differing beliefs about the normal prognosis of the problem. Those who believed that BPH typically proceeds eventually to blockage of the urethra favoured early surgical intervention. Others recommended “watchful waiting,” believing that many cases would never require surgery. Early intervention would lead to much unnecessary surgery, with a significant rate of serious side effects.

Of course physicians’ patterns of practice depend on their beliefs about the relative benefits and risks of particular interventions. Would one want them to behave other-

wise? But the observed variations in practice indicate that these beliefs are highly variable from one clinician to another, and some of them (at least) are wrong. In principle, and often in practice, empirical evidence can be brought to bear to determine which is which.

In the case of BPH, the evidence turned out to support watchful waiting. Moreover patients, when given information about risks and benefits, tended strongly to favour watchful waiting. But until the question was taken up as a research program by the Dartmouth investigators, the alternative beliefs were never tested. Individual surgeons just went ahead doing what they thought best – like the obstetricians in Boston.

Plus ça change. Berenson and colleagues (2009: 937) have studied the diffusion of (expensive) telemedicine technology in American intensive care units (eICU):

We explore the reasons hospitals chose to adopt or reject an innovative telemedicine approach ... Hospital clinical leaders hold strong views but have little objective information on which to judge the worthiness of this innovation.

Ignorance is strength?

The BPH and tonsillectomy studies were important because they each provided a response to the standard defensive “yabbut”: “Who knows which rate is right?” In these cases, the high rates of surgery were the wrong ones. The general blocking tactic follows one of two arguments. One is to assert that low-use populations are, or may be, underserved. Their access is being limited by shortages of personnel or equipment, or inability to pay – or something. The other is that “everything is beautiful in its own way.” Patients’ needs differ, so patterns of care vary because knowledgeable and responsible clinicians provide the care appropriate to those differing needs. End of story.

The first argument emerged in the 1950s in response to observations that hospital utilization rates were much lower in pre-paid group practices than in the general fee-for-service community. It largely fell out of favour after controlled trials that randomly assigned patients to pre-paid group practice or community care demonstrated that organizational settings account for the differences in use, and that low users were not underserved.

The second argument in effect denies that variations represent a problem. It places the burden of proof on those who would suggest otherwise (see the remarks of Senator John Kerry as reported by Pear [2009]). This response has worked for decades, but a great deal of progress has been made in the last quarter-century, as reflected in Wennberg’s recognition by *Health Affairs*. A remarkable pair of papers by Fisher and colleagues (2003a,b) show very large regional differences in service utilization and expenditures by Medicare beneficiaries (ages > 65) in the United States, after standardizing for measures of patient health status. And high use and cost areas have higher mortality rates, though equivalent levels of patient satisfaction. More is not better; it’s worse.³

Important information lies behind the aggregates. The researchers have catego-

alized specific services as (1) effective care, (2) preference-sensitive care and (3) supply-sensitive care.

The first are services or procedures supported by clinical evidence as improving the health of patients. No trade-offs are involved – do it! The second are those interventions for which there is a balance of risks and benefits, and patients' values and preferences should govern the choice. The third are those whose utilization is strongly associated with the local availability of resources – personnel, equipment and facilities. One might think of these three categories as medically driven, patient driven and capacity driven.

Interregional variations in use and cost reflect variations in supply-sensitive services – full stop. This is not to say that differences in patients' needs or preferences play no role in influencing utilization. But these factors wash out in aggregate. The large regional variations in average rates of utilization and cost are driven from the supply side, by differences in clinicians' choices, not in patient needs or preferences.

Up pops another standard yabbut – what about “quality of care”? Could more servicing have benefits that are not captured by mortality or patient satisfaction? The Dartmouth investigators have approached this question indirectly, showing large variations in servicing and costs among academic medical centres that are generally acknowledged to provide care of the highest standard. The first study compared Boston and New Haven (Harvard and Yale); more recent papers have expanded the number of centres included.

Boston was, on average, much more expensive than New Haven in caring for Medicare patients. Twenty years later, the Mayo and Cleveland clinics turn out to be much less costly than Johns Hopkins or UCLA. Uwe Reinhardt has quipped that in the United States, “the finest medical care in the world costs twice as much as the finest medical care in the world.” There's no reason for it, it's just our policy.

It is tempting to describe these differences as “cost without benefit,” but that would be misleading. All costs benefit someone; that is why, when the Gawande story broke, Senator Kerry was so quick to dismiss the regional variations findings (Pear 2009). He showed no obvious competence; his comments would be easily recognized by Gawande's Texas surgeon. But Senator Kerry has a very clear understanding that billions of federal dollars flow into his state as income for its highly developed medical-industrial complex. Serious attention to expenditure variations would threaten those incomes. He is instantly on the attack.

The United States will spend approximately \$2.4 trillion on health services this year, and every dollar flows into someone's pocket. Their representatives, political and professional, stand on guard to make sure the money keeps coming – \$2.4 trillion pays for some very heavy artillery indeed. “Who ever knew Truth put to the worse, in a free and open encounter?” asked Milton. “Who ever saw a free and open encounter?” replied Satan. We are certainly not seeing one now.

The accounting identity linking total expenditures and total incomes is the most

fundamental contribution that economic analysis makes to the understanding of real-world health systems. It provides the primary explanation for 40 years of political indifference to the variations data. If President Obama can “fix” McAllen, or anywhere else, some incomes will have to be cut. But if not... not.

Beyond that powerful insight, the quality of economic contributions becomes much more uneven. Economists are not, in the main, stupid,⁴ but they have said some remarkably stupid things about health. The assumption that all health services utilization follows from the decisions of more or less informed “consumers,” for example, implies that clinical variations must result from regional differences in “consumer tastes.” The residents of McAllen simply have a particularly intense taste for various forms of health services, just as they might have a particular taste for chocolate ice cream. There is nothing to “fix”; *de gustibus non est disputandum*.

This is an essentially theological position, as impervious to fact or argument as “creation science.”⁵ It parallels the medical claim that clinical variations simply reflect clinicians’ appropriate responses to differing patient needs. Both are circular arguments, positing an inherently unobservable concept – tastes, or needs – whose variations are inferred from observed variations in use and then serve to justify those variations. If direct observations fail to confirm belief, the observations are wrong.

Another distraction is provided by the common economic fascination with trade-offs. This argument emerges in the mindless mantra that no system can simultaneously achieve universal coverage, high-quality care and cost control. Its roots lie in the original fallacy that “more is better” and that cost equals quality. Its political appeal may be that it appears to justify the floundering of American health policy. Clinical variations provide a direct refutation (as, for that matter, does international experience); the mantra is simply false. But economists, even some health economists, have been slow to absorb that message.

Many were quick, however, to absorb the message that patients served by pre-paid group practices, later health maintenance organizations, made systematically less use of hospitals and generated significantly lower costs. These observations could be interpreted in a standard framework of economic motivations and incentives – contrasting capitation with fee-for-service payment. The obvious implication was that a “world of competing HMOs” would curb cost escalation and could offer better-quality care. Roll on the Managed Care Revolution! (How can we get it into Canada?) Economists (including this one) failed to reflect carefully on the implications of clinical variations.

That physicians have powerful economic motives and respond to economic incentives is hardly a debatable proposition. But the variations emerge, then and now, within a relatively homogeneous reimbursement environment. It is true that much of the regional variation is correlated with variations in capacity, personnel and equipment. But much is not, and in any case capacity is not exogenous. It responds to clinicians’ views as to what is needed.

The “surgical signature” underlined the importance of physicians’ individual preferences for, or confidence in, particular patterns of intervention.⁶ The clustering of behaviour also indicates strongly that physicians’ preferences are formed within, and respond to, a local culture. In the mid-1960s, when normal deliveries stayed five days in Boston and three in San Francisco, physicians’ economic motivations were as irrelevant as patients’ needs.

In short, economists’ “explanations” of patterns of utilization, and the physician behaviour that drives them, suffered from the characteristic flaws of economic reasoning. The assumptions of the representative agent – *the physician*, analogous to *the consumer* or *the firm* – leads to

In short, economists’ “explanations” of patterns of utilization, and the physician behaviour that drives them, suffered from the characteristic flaws of economic reasoning.

a focus on aggregates that suppresses the behavioural information in variations data. This, in turn, encourages oversimplification of the objectives postulated for physicians, and the strategies available to them. We

impose *a priori* far too narrow a view both of what physicians are trying to do, and of how they go about doing it – not necessarily wrong, but seriously incomplete. The variations literature shows what we have been missing.

Finally, there is a long-standing tradition of such work in Canada as well, notably the early work of Eugene Vayda and colleagues (1976; Stockwell and Vayda 1979) and the continuing work of Leslie and Noralou Roos and theirs (1977, 1983, 1986). More recently, Alter and colleagues (2008: 187) report that in Ontario

[r]egional per capita cardiologist supply varied more than twofold across regions, but was inversely related to the regional cardiovascular disease burden. ... Residents in areas with more cardiologists were more likely to receive some form of cardiac intervention. ... However, the intensity of provision of cardiac health services was unrelated to regional cardiovascular disease burden and was not associated with improved survival.

In short, capacity-driven utilization.

The monumental *Canadian Cardiac Atlas* (Tu et al. 2006) includes a study of hospital admission rates for leading cardiac diagnoses (Hall and Tu 2003). The authors found very high interregional variations, with gradients rising strongly from west to east, and from large cities to rural areas. The Canadian average admission rate was just under double the rate in the city of Vancouver, and the discrepancy in patient days was even larger.⁷ The authors comment, with some understatement: “There is consid-

erable regional variation in the cardiovascular hospitalization rates across the country that may be amenable to further interventional strategies” (Hall and Tu 2003: 1123).

Yet again, much has been made in the professional and public rhetoric of the inadequacy of CT and MRI capacity in Canada, and very large amounts of money have been allocated to a rapid expansion and modernization of diagnostic imaging

These huge international variations in imaging availability are unconnected with any evidence of differences in patient needs or outcomes.

facilities. The survey by the Canadian Institute for Health Information, *Medical Imaging in Canada: 2007* (CIHI 2008) documents the corresponding rapid increase in capacity for, utilization of and expen-

ditures on these procedures. But it also documents the wide interprovincial variations in capacity and use and, more importantly, the extraordinary international variations.

Japan had 92.6 CT scanners and 40.1 MRI machines per million population in 2005; the Netherlands had 5.8 and 5.6 (CIHI 2008, figures 39 and 40). The United States had 45.3 and 26.6; Germany had 15.4 and 7.1. Canada, at 12.1 and 6.1 (in 2006), was just below the medians of 14.7 and 6.9. But there is no “international standard”; country rates are all over the map and averages mean nothing. In these circumstances, to try to “keep up with the rest of the world” is to chase a chimera. There is no “rest of the world” in any meaningful sense.

These huge international variations in imaging availability are unconnected with any evidence of differences in patient needs or outcomes. Yet diagnostic imaging is, along with laboratory testing and pharmaceuticals, one of the primary sources of cost escalation in Canada. A focus on these sectors might be more productive than general blather about “sustainability.”

The implications of these Canadian reports, fragmentary as they are, are straightforward. Clinical variations, driven by physician preferences and local medical cultures, not by patient needs and evidence of effectiveness, are a major issue in Canada as well. They have not been as intensively studied as in the United States, but they have been studied, they have been found and they are large. The significance of such variations has finally penetrated the highest political levels in the United States, although that country’s bizarre political system may be incapable of reacting sensibly. In Canada, they are not even on the radar.

“Only in America, you say? Pity.”

Notes

¹ Miami is higher, but has much higher labour and living costs.

- ² Large regional variations do *not* imply that surgical procedures, or medical services generally, are simply distributed capriciously. Research supports the obvious; care tends to go where it is needed. Health system is mostly used by sick people and sicker people use more care – and women in both Boston and San Francisco were giving birth. But, following Rosé's Law, variations in *population* rates are not explained by variations in needs.
- ³ The Dartmouth *oeuvre* is now huge, and referencing quickly becomes unwieldy. Key findings are however collected together, with supporting references, in Fisher (2007).
- ⁴ Some are – names not available on request – and a few are simply “on the take”.
- ⁵ Persistent nonsense is often rooted in economic interests. The “consumer tastes” fantasy supports various schemes such as Medical Savings Accounts, or “Consumer-Directed Health Care” that would transfer costs from taxpayers to patients – i.e. from the healthy and wealthy to the unhealthy and unwealthy – while improving access for the wealthy and unhealthy. The naked redistributive agenda is obscured by “econofog” (a very thick economist).
- ⁶ Evidence eventually matters; tonsillectomies are rarely done today because the believers have died.
- ⁷ Their data are from the late 1990s, but there is no reason to expect that these differentials have changed.

REFERENCES

- Alter, D.A., T.A. Stukel and A. Newman. 2008. “The Relationship between Physician Supply, Cardiovascular Health Service Use and Cardiac Disease Burden in Ontario: Supply–Need Mismatch.” *Canadian Journal of Cardiology* 24(3): 187–93.
- Berenson, R.A., J.M. Grossman and E.A. November. 2009 (August 20). “Does Telemonitoring of Patients – the eICU – Improve Intensive Care?” *Health Affairs* 28(5): w937-w947. doi: 10.1377/hlthaff.28.5.w937.
- Canadian Institute for Health Information (CIHI). 2008. *Medical Imaging in Canada: 2007*. Ottawa: Author.
- Fisher, E.S. 2007 (May 24). “Pay-for-Performance: More Than Rearranging the Deck Chairs?” Robert and Alma Moreton Lecture, Center for the Evaluative Clinical Sciences, Dartmouth Medical School. PowerPoint presentation.
- Fisher, E.S., D.E. Wennberg, T.A. Stukel, D.J. Gottlieb, F.L. Lucas and E.L. Pinder. 2003a. “The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care.” *Annals of Internal Medicine* 138(4): 273–87.
- Fisher, E.S., D.E. Wennberg, T.A. Stukel, D.J. Gottlieb, F.L. Lucas and E.L. Pinder. 2003b. “The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care.” *Annals of Internal Medicine* 138(4): 288–98.
- Gawande, A. 2009 (June 1). “The Cost Conundrum: What a Texas Town Can Teach Us about Health Care.” *The New Yorker*. Retrieved September 17, 2009. <http://www.newyorker.com/reporting/2009/06/01/090601fa_fact_gawande?currentPage=all>.
- Hall, R.E. and J.V. Tu. 2003. “Hospitalization Rates and Length of Stay for Cardiovascular Conditions in Canada, 1994 to 1999.” *Canadian Journal of Cardiology* 19(10): 1123–31.

- Institute of Medicine of the National Academies. 2009 (May 11). "2008 Lienhard Award Recipient: John E. Wennberg." Retrieved September 17, 2009. <<http://www.iom.edu/?id=59003>>.
- Orszag, P.R. 2008 (February). *Geographic Variation in Health Care Spending*. Washington, DC: Congressional Budget Office, Congress of the United States.
- Orszag, P.R. and P. Ellis. 2007 (November 1). "The Challenge of Rising Health Care Costs – A View from the Congressional Budget Office." *New England Journal of Medicine* 357(18): 1793–95.
- Pear, R. 2009 (June 8). "Health Care Spending Disparities Stir a Fight." *The New York Times*. Retrieved September 17, 2009. <<http://www.nytimes.com/2009/06/09/us/politics/09health.html>>.
- Roos, L.L. 1983 (April). "Supply, Workload and Utilization: A Population-Based Analysis of Surgery in Rural Manitoba." *American Journal of Public Health* 73(4): 414–21.
- Roos, N.P., G. Flowerdew, A. Wajda and R.B. Tate. 1986 (January). "Variations in Physicians' Hospitalization Practices: A Population-Based Study in Manitoba, Canada." *American Journal of Public Health* 76(1): 45–51.
- Roos, N.P., L.L. Roos and P.D. Henteleff. 1977 (August 18). "Elective Surgical Rates – Do High Rates Mean Lower Standards? Tonsillectomy and Adenoidectomy in Manitoba." *New England Journal of Medicine* 297(7): 360–65.
- Stockwell, H. and E. Vayda. 1979 (April). "Variations in Surgery in Ontario." *Medical Care* 17(4): 390–96.
- Tu, J.V., W.A. Ghali, L. Pilote and S. Brien, eds. 2006. *Canadian Cardiovascular Atlas*. Toronto: Canadian Cardiac Outcomes Research Team, Institute for Clinical Evaluative Sciences (ICES).
- Vayda, E., M. Morrison and G.D. Anderson. 1976 (May). "Surgical Rates in the Canadian Provinces, 1968–1972." *Canadian Journal of Surgery* 19(3): 235–42.
- Wennberg J. and A. Gittelsohn. 1973. "Small Area Variations in Health Care Delivery." *Science* 182 (117): 1102-8.
- Wennberg, J. and A. Gittelsohn. 1982 (April). "Variations in Medical Care among Small Areas." *Scientific American* 246(4): 120–34.