The Dartmouth Team Responds – Again -- to Ms. Abelson and Mr. Harris of the *New York Times*

Jonathan Skinner, PhD
Elliott S. Fisher, MD MPH
June 22, 2010

We were pleased to see that Reed Abelson and Gardiner Harris, the authors of the June 4th New York Times article critical of the Dartmouth Atlas and research, have acknowledged our concerns and clarified the record in their posting on the New York Times webpage. But we remain disappointed by their recent attempt to defend their article.

Much of the debate has now ended up firmly in the statistical weeds where only the hardy dare venture. But a careful reading of their more recent article reveals several points of clarification (and contention) that we think are important to understand.

1. Distinguishing between the Atlas Maps/Data and the Peer-Reviewed Research. Abelson and Harris now better understand the fundamental difference between the Dartmouth Atlas data and the Dartmouth research published in peer-reviewed journals. Their fear is that Congress and the Administration have relied (and will rely) primarily on the Atlas data (particularly measures of regional and hospital measures that do not always control for differences in illness or local prices) to penalize or reward individual hospitals or regions. We certainly share this concern.

But our fundamental message – starting with John E. Wennberg's work in the 1970s -- is twofold: (a) there are wide variations in health care utilization across regions, and (b) more spending does not necessarily result in better health care quality, access to care, or health outcomes. The gold-standard measures of regional utilization and health outcomes come not from maps, but instead from the multiple datasets which allow us to adjust for differences across regions in health status, prices, poverty, and other factors, and which have been published in many peer-reviewed journals such as the *New England Journal of Medicine*.[1] And even after such adjustments, we find that some well-organized health care systems can and do provide high quality care for low cost – which in turn is suggestive of the vast amount of waste in the U.S. health care system.[2]

We are perhaps more optimistic than Abelson and Harris that policymakers would rely on more reliable measures for designing new incentive systems under health care reform than to use

[1] The datasets used in the research are large by most standards, with 15,000 or more people, but once one divides up those 15,000 into the 306 hospital referral regions (HRRs), statistical precision suffers too much to create maps.

[2] Ms. Abelson and Mr. Harris are shocked that a leading health care academic is unaware that some of the Dartmouth maps show Medicare expenditures that are unadjusted for differences in prices. But this simply emphasizes our point – that the first stop for academics is the research, and not necessarily the Atlas data.

unadjusted spending measures. But we think that reporting the clean per capita age-sex-race adjusted Medicare expenditure measures is worthwhile, despite the risk that a politician may misinterpret them.

So what can be learned from the Atlas website? The first is that there are wide and systematic differences in a wide range of spending and utilization measures across regions, and even across hospitals within a given region. Inexplicably, the New York Times reporters entirely ignore the large number of measures such as back surgery rates, hospital days, and other measures that reflect true utilization rates. Indeed, a landmark article by Ms. Abelson used the Dartmouth Atlas data on very high rates of cardiac stents in Elyria, Ohio – a finding that was confirmed by her own investigative research.[3]

The hospital-level data is also very useful in comparing utilization and spending rates across comparable institutions – for example, why do UCLA and Cedars-Sinai provide so much more (price-adjusted) intensive care than comparable research medical centers like the University of Chicago, located in one of the poorest neighborhoods in Chicago? Which are the hospitals that manage to have world-class quality reputations (based for example on the US News and World Report hospital rankings) but do so with low costs? Without the Atlas measures, these questions could not be asked.

Are the measures perfect? Of course not. But they are a first step. The cost and utilization measures provide important feedback to providers and a building block for the next generation of measures that can be used to reward improved performance.

2. Quality Measures in the Dartmouth Atlas. The original New York Times article stated that

Measures of quality of care are not part of the formula. For all anyone knows, patients could be dying in far greater numbers in hospitals in the beige regions than hospitals in the brown ones, and Dartmouth's maps would not pick up that difference.

As Abelson and Harris now acknowledge, there are a variety of quality measures by region reported in the Atlas: quality of care for diabetic patients, and (had they clicked a little further) rates of amputations for diabetics, CMS hospital quality measures, and patient satisfaction measures with their hospital care (pain control, noisiness, cleanliness of rooms, etc) .

They claimed that we didn't report whether patients are "dying in far greater numbers in hospitals" and in response we pointed out that we do measure this – the percentage of people who die in a hospital. (Given the strong desire of many people to die at home or in a hospice,

[3] Another example of Atlas data that was confirmed by field investigation was Dr. Atul Gawande's celebrated study of McAllen and El Paso. He could use unadjusted spending and utilization measures for the two regions because prices were similar in the two regions, as were objective measures of health status such as cardiovascular or cancer deaths. And the nearly five-fold differences in home health care spending per capita were unlikely to be explained by anything other than a "culture of entrepreneurship."

this is indeed a measure that potential patients should think about before seeking care at a specific hospital.)[4] In our initial response we focused on regional measures of quality. Apparently Abelson and Harris were more interested in hospital level data – but once again, the Atlas website provides the Hospital Compare quality measures from CMS (reported in the 2008 Atlas here, Appendix 3h and on the website.)

In sum, their original statement that "measures of quality of care are not part of the formula" is incorrect. They now acknowledge that the Atlas measures (as well as the research) do pay attention to quality. They just don't *like* our quality measures. But the measures reported by the Atlas are drawn from nationally endorsed sources – NCQA and CMS. And Dr. Fisher co-chaired a National Quality Forum committee to accelerate the adoption and implementation of more meaningful measures across the country. The measures are not perfect, because they do not capture the full spectrum of health care. *But they're better than the alternative, which is total ignorance.*

They also now acknowledge that price-adjusted atlas spending measures are indeed available, but difficult to find. We concede the point and are working to improve the web interface and make these data generally available to researchers and policy makers for multiple years.[5]

3. Is More Worse or Just Not Better? Abelson and Harris are not content with (appropriately) worrying about the misuse of the Atlas data in the public policy arena. But they also take on the published research, and in particular the two 2003 *Annals of Internal Medicine* articles. Indeed, in a posting here from The Health Care Blog, Mr. Harris asserts:

In an aside, when was the last time you saw researchers so profoundly mischaracterize their own work? How is it possible that they could claim their annals pieces concluded something when they didn't? I can't remember ever seeing that happen.

We are deeply disappointed by this level of discourse -- a personal attack on our integrity. And we can understand that Mr. Harris may have misunderstood some of the statistical nuances of the research. For example, Ms. Abelson and Mr. Harris state that:

In statistical terms, this claim is referred to as a negative correlation between spending and health outcomes, which means that when spending goes up, the health of patients goes down.

[4] They also fault us for not combining the cost and quality measures on a single map. Unfortunately there is no clear way at this point to combine the two measures.

[5] Before posting on the website, we also want to ensure that our newly developed price-adjustment methods are consistent with the price-adjustment methods in use by MedPAC and other federal agencies. Price adjustment is not simple – it took us 2 years to get it right.

This is an incorrect statement. They have confused the idea of a correlation (high spending hospitals tend to do poorly on most measures of quality and outcomes) with causation (if a hospital spends more money, outcomes for those patients will get worse).[6]

The more fundamental point, however, is their claim that we have misrepresented the findings of the two 2003 *Annals of Internal Medicine* studies. They state that

The Dartmouth work has long been cited as proving that regions and hospitals that spend less on health care provide better care than regions and hospitals that spend more.... As the article noted, [Dr. Fisher] asked in Congressional testimony last year, "Why are access and quality worse in high-spending regions?"

But now we come to their smoking gun(s):

Those [Annals] studies did conclude that there was no association between higher spending and better health, but they did not show any link between higher spending and worse health.

One of the paper's arguments was summarized in the abstract this way: "Neither quality of care nor access to care appear to be better for Medicare enrollees in higher-spending regions."

The second paper's summary: "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."

Ms. Abelson and Mr. Harris are correct to note that the studies conclusively rule out the hypothesis that more spending is associated with better outcomes. The major aim of the paper was to examine empirically the assumption that spending more leads to better outcomes. And the paper does show, conclusively, that more care does not necessarily result in better care.

But anyone who reads both articles will come away with more than that – a finding that care was generally worse in high spending regions across all dimensions: technical quality (e.g. the right medication for a heart attack patient), access to care (e.g. did patients have timely access to a physician after discharge) and health outcome measures (e.g. survival). We did a quick tabulation of the quality, access and outcome measures in the articles – a total of 42 different measures. Of the total of 42, 23 showed significantly worse outcomes in high-spending regions, 14 showed no significant effects, and just 5 showed significant positive effects in high-spending regions.[7] And counting in this way doesn't weight the different measures – for example, one of

[6] Thus the correlation between the population of storks and birth rates does not mean that storks are causal in ensuring that the baby is delivered to its mother. We're not making this up – see for example this study.

[7] In order of the tables (and starting with Part 1): **Those showing significantly worse care in high cost regions**: MI: reperfusion, aspirin (hospital), aspirin (discharge), ACE inhibitors; MCBS: influenza, pneumonia, pap; Access: angiography, angiography (appropriate patients), PCI, stress tests, MD visit within 30 days, re-admission within 30 days (a higher rate is a marker for poorer quality); MCBS: usual source of care, waited more than 30 minutes for ER

the 23 worse outcomes is that heart attack patients are more likely to die in high cost regions, while one of the 5 better outcomes is that patients with heart attacks are more likely to be treated with bypass surgery – not such a clearcut "quality" measure as not every AMI patient will benefit from bypass surgery and there are real risks of overtreatment.

We construct an index of quality: the ratio of significantly worse measures to significantly better measures in high cost regions. Our ratio is a whopping 4.8. In other words, there are nearly 5 measures showing quality of care in high cost regions to be worse for every measure that suggests high cost regions provide better care.

So when Abelson and Harris claim that Fisher and others are overstating the results of the paper, they are simply wrong. We're not sure why they are pursuing this, unless it reflects a lack of experience in reading and interpreting scientific papers. But that is no excuse for making unfounded accusations about our scientific integrity.

As well, there are a number of other papers (described in our longer essay provided to the *New York Times*) that shows the same negative association between spending and outcomes.

We have also written in more detail in this longer essay, available also on the Times website, about a newer set of studies by other researchers that are consistent with our central premise that it doesn't matter how much you spend, it's how you spend the money that matters. So with end-of-life spending on "supply-sensitive" care such as discretionary hospital and ICU days, feeding tubes, and mechanical ventilation, it's not so surprising that there is a negative association between spending and outcomes. But when the spending is for providing neonatologists in understaffed regions of the country (as shown by David Goodman's study), better diagnosis and treatment of tourists treated for MI in Florida (here) or premature babies, or for Amber Barnato's study of ICU patients with designated protocols, then it makes sense to find that dollars could translate into better outcomes.[8]

---

visits , outpatient visits, and MD visits, (more is worse for these three measures), did not see an MD for a problem; (Part 2) Colorectal Cancer survival, MI survival, (acute care expenditure measure): global quality of medical care. **Non-significant (zero) effects.** Part I: MI: Beta blocker at discharge, mammography; MCBS ER visit, trouble getting care, delayed care because of costs. Part II: Hip fracture survival, MCBS survival, functional status, (Acute care expenditure index) hip fracture survival, colorectal cancer survival, MCBS survival, access to care, information giving, technical skills. **Better care in high cost regions**: Part 1: Beta blocker at hospital, CABG post-MI, whether one sees a cardiologist or surgeon post MI, interpersonal skills of physician. Note: there is no presumption that seeing a family or general practitioner, internal medicine, or other specialist is intrinsically better or worse for quality of care, thus these are not ranked. We did give the nod, however, to seeing a cardiologist or surgeon.

[8] No matter whether the association between spending and outcomes is positive or negative, typically the association is quite weak – in other words, it looks like a scatter shot for hospitals or regions when it comes to a graph of quality and outcomes (see our paper that makes this point in greater detail). Abelson and Harris try to claim that we are being inconsistent, for example when they write that "*In a recent article in The New England Journal of Medicine, for example, Mr. Skinner wrote that "empirical data from a variety of studies shows little or no association between spending and health outcomes."* This is of course entirely consistent with the

4. Fessing Up to One's (Minor) Mistakes. In the original New York Times Article, Abelson and Harris stated:

A 2003 study found that patients who lived in places most expensive for the Medicare program received no better care than those who lived in cheaper areas.

Because some regions spent nearly a third more than other regions without any apparent benefit, the Dartmouth team concluded that at least one dollar in three was wasted by Medicare.

In our June 7th response, we noted that the 2003 study stated that "that up to 30%" savings could be achieved, rather than at least one-third. As we noted then, this is a minor "nitpick." What they could have said was "nitpicking point taken – but you have elsewhere pointed to much larger potential savings of as much as 42% from other studies." And we would have happily agreed with that statement. Instead, they refuse to acknowledge their error.[9]

5. The Challenges of Risk Adjustment. Our NEJM article on regional differences in risk adjustment does present a challenge to current risk adjustment methods and highlights the need for more work. But we disagree that the finding undermines the variations work. The key point is that we find a systematic bias, not a random one – high intensity regions tend to over-diagnose relative to low intensity regions. This makes it even more surprising that when we use even the current flawed methods, marked variations remain. Indeed, the biases go against us finding zero or negative associations between intensity and outcomes: risk-adjusted outcomes will look much better for high intensity regions, and their risk-adjusted costs will be attenuated. And it also reinforces our approach to present the clean data and not try to over-adjust, because sometimes the cure is worse than the disease.

One of the reasons why we favor using cohorts of patients -- e.g. those with heart attacks or a hip fracture, or those who are in their last six months or two years of life – is precisely because the biases arising from provider-driven risk adjustment are less serious in this group. A hip fracture patient is going to be pretty sick, no matter how much or how little risk adjustment is done. And when we do use these types of cohorts, we find remarkable stability in our hospital cost measures, as is shown here. [10]

---

typically faint negative association between spending and outcomes that we find in the Annals and other studies – the slope is significant, but the explained variation (or the R^2) is trivial, which means in statistical terms "little" association.

[10] Amitabh Chandra, a professor at the Kennedy School of Government, was quoted by Ms. Abelson and Mr. Harris as claiming that our risk adjustment paper "threatened the validity of the last 20 years of their research." But Professor Chandra emailed Mr. Harris back in May providing much the same explanation as ours. This is yet another example of an expert quoted by Ms. Abelson and Mr. Harris whose quotation is technically correct, but who did not agree with the interpretation placed on that quotation. See Maggie Mahar's blog posting for more discussion of this point.

To sum up, we can all agree that improving these measures should be a first priority for policy makers and academics. And we also acknowledge the limitations of the research, and how much there is that we don't know. Indeed, Henry Aaron, a respected health economist at the Brookings Institution, wrote about the issues and concerns the New York Times article should have addressed but didn't. We would have welcomed such a thoughtful and balanced approach.

Perhaps we are old-fashioned, but what we find most puzzling at the end of the day is why this process should have been so adversarial. It began in the winter with Mr. Harris announcing at the beginning of an interview that he was going to "take down" the Dartmouth Atlas (as documented in Maggie Mahar's blog). And it is ending (we hope) with Mr. Harris's posting on The Health Care Blog (quoted above) questioning our ethical standards. This saddens us because of the missed opportunity to improve the dialogue in Washington and elsewhere about the strengths and limitations of research on regional variation.