In a recent paper, Reschovsky and colleagues claimed that nearly all regional variations in health care expenditures in the United States can be explained by differences in health status across regions. Their results are thus quite different from many other studies documenting real geographic variations, not just in the United States, but around the world (Skinner, 2012). In these comments, I suggest three reasons why their results are so much different from other studies.

The first and most important is a peculiarity of commonly used risk adjustment methods. Hierarchical condition categories, or HCCs, are based on billing data, and designed to “adjust” for the fact that some people are in fact sicker than others. But increasingly, we have begun to realize that HCCs are severely flawed, because physicians and hospitals cannot bill Medicare without a diagnosis. So for example if an individual in McAllen Texas is given a stent by an aggressive cardiologist – as detailed in Atul Gawande’s New Yorker article – she is coded as having serious heart disease. By contrast, that same patient in Grand Junction, Colorado is sent home with at best a much less serious condition (if at all). In this case, the authors’ use of the HCC billing codes would falsely “explain” the more aggressive cardiologist’s behavior as the consequence of poor health status, rather than attributing it correctly to intensive physician behavior.

As we have shown in previous research published in the New England Journal of Medicine, JAMA, and BMJ, these biases are severe and will lead to highly misleading conclusions. Reschovsky and colleagues attempt to deal with the problem in a variety of ways. For example, they consolidate 5 types of diabetes diagnoses into one, but that does nothing to address the fundamental problem that in high-intensity regions, there is a lower threshold for the diagnosis of diabetes, leading to more individuals with that diagnosis. This leads naturally to the “reverse causation” problem, that intensive physician treatment leads to greater rates of disease diagnosis.

Reschovsky et al. (2013) have also claimed that end-of-life measures used by the Dartmouth Atlas are not adjusted for health risk. But their use of the HCC risk-adjusters for end-of-life measures leads to exactly the same biases as noted above. What happens when end-of-life spending is risk-adjusted in a sensible way? Amy Kelley, a physician at Mt. Sinai Hospital in New York, has done precisely this. She and colleagues examined actual Medicare expenditures in the last six months of life for 2,394 individuals in the Health and Retirement Study. The study, published in the Annals of Internal Medicine, included state-of-the-art risk adjusters, such as
self-reported health, activities of daily living, wealth, education, functional status, and type of disease. She found that even when these risk adjusters were included, the observed wide variation in end-of-life spending was essentially unaffected – thus directly contradicting the Reschovsky et al. (2013) finding.

Another approach that Reschovsky et al. (2013) use to address the reverse-causation problem is to focus on “non-discretionary” measures of disease, like hip fractures, which they find are considerably higher in high-cost regions. Wouldn’t that lend credence to their claim that poor health explains higher spending? I believe the answer is still no – because we cannot reproduce even those basic findings when we use our population-based Medicare claims data. For example, in their study, enrollees in the highest-spending quintile are 73% more likely to report hip fractures/dislocations. In our Medicare data, the equivalent ratio for hip fractures is just 13.5%. They also find much larger differences in mortality rates between the lowest and highest quintiles, again on the order of three times what we find. (When I tried to match their smaller sample of sites to our HRR-level data, I found no difference at all in mortality rates between high and low-expenditure regions.)

Why the divergence? I’m not entirely sure, but I suspect that it’s because their sample is based on physicians, not patients. In high-cost regions, patients, particularly sick ones, see many different physicians, often specialists. Thus a sample of physicians in a high-cost region will unearth sicker patients, leading to a sample design with built-in biases. They acknowledge this problem and to their credit have attempted to reweight the data to avoid such biases. But I am not convinced that the reweighting fixes the fundamental flaws of their original sample, given that their sample does not match even the basic empirical patterns observed in the nationally representative Medicare claims data.

The final concern with this study is a more technical one, so bear with me. Typically, HCC risk adjustment is predetermined by the HCC formula designed by CMS. If someone has diagnosed COPD, Medicare would implicitly allow (say) 23% more expenditures as a consequence. As far as I can tell from the methods section, this is not what they do. Instead, they throw the HCC diagnostic codes into the regression, and let the regression do its best to explain the regional variation. Regressions with dozens of explanatory variables are designed to “soak up” as much variation as possible -- regardless of clinical plausibility. That they alone use this approach is perhaps another reason why they systematically attribute too much importance to HCC measures.

It is informative to consider the interim report by the IOM-convened panel of experts on regional variations – a panel that did not include any Dartmouth faculty. The panel, chaired by Harvard professor Joseph Newhouse and Harvard provost Alan Garber, sorted through the existing evidence, commissioned new studies, and even ran their own analyses. They concluded that “Although a non-trivial amount of geographic variation can be explained by specific demographic and, potentially, health status variables, a substantial amount of variation remains unexplained.” (p. 16) Nothing I have seen in Reschovsky et al. (2013) would lead me to revise the IOM panel’s conclusion.