Review of Desai and McWilliams, "Consequences of the 340B Drug Pricing Program", *New England Journal of Medicine*, 2018

This paper uses data from Medicare claims and a regression discontinuity design to estimate the effects of the 340B Drug Pricing program on hospital-level physician count and administration of parenteral drugs outcomes and patient-level receipt of care and mortality. I have a number of concerns with the empirical strategy that can be categorized into three broad groups. First, I report concerns with implementation of the regression discontinuity strategy in the context of hospital level outcomes. Next, I report some additional concerns with implementation of the regression discontinuity strategy in the context of beneficiary level outcomes. Finally, I raise some concerns about the interpretation of the results as reported.

Hospital level analysis

1. There may be a fundamental error in the interpretation of the estimates reported in the paper and in the supplementary analyses. The authors write the regression equation as

$$E(Y) = \beta_0 + \beta_1 \text{Eligible} + \beta_2 \text{DSH} + \beta_3 \text{Eligible} \times \text{DSH} + \beta_4 \mathbf{X} + u \quad (1)$$

where Eligible $=$ 1 if DSH $>$ 11.75. Then they write that the estimate of β_1 is the treatment effect. In this equation, β_1 is the effect of switching from being non-eligible to being eligible *for a hypothetical hospital with a DSH percentage equal to 0.* Clearly, this hospital is outside the scope of the sample, and is not a hypothetical hospital of any policy interest. Instead, the regression discontinuity regression should have been specified as follows, with the DSH variable specified as its deviation from the threshold value of 11.75.

$$E(Y) = \beta_0 + \beta_1 \text{Eligible} + \beta_2 (\text{DSH} - 11.75) + \beta_3 \text{Eligible} \times (\text{DSH} - 11.75) + \beta_4 \mathbf{X} + u \quad (2)$$

In this specification, β_1 is the effect of switching from being non-eligible to being eligible *for a hypothetical hospital with a DSH of 11.75.* This is the estimate of policy interest because it shows what would happen to a hospital close to the discontinuity if a policy were changed, i.e., for a hospital on the margin. I cannot determine, from the paper and the supplementary materials document, whether this is an editorial error or a substantive error. If it is substantive, i.e., the authors estimated equation (1) and report the estimates of β_1, then all the reported regression results cannot be interpreted as the treatment effects from a regression discontinuity regression.

2. The authors drop all hospitals with less than 50 beds, which they state include critical access hospitals and sole community hospitals. The argument that critical access hospitals and sole community hospitals are different from other short term general hospitals is a reasonable one. But there are plenty of small short term general hospitals and it is not clear why they should be dropped from the analysis. Admittedly, such hospitals would get small weights in the beds weighted regressions, but they do constitute a substantial fraction of hospitals. Their inclusion could change results substantively.

3. The authors also drop hospitals with DSH percentages within 1 percentage point of the threshold, leading to omission of a non-trivial number of hospitals. Omission of these hospitals quite likely has substantive implications for the results. The authors argue for dropping hospitals within 1 percentage point as being due to misclassification, which is an issue but a common one in studies that use regression discontinuity designs. It can be dealt with, as the authors do, using instrumental variables methods. But for the main analysis, the authors drop these observations and cite a paper that deals with heaping, which is not the same thing as misclassification. More substantively, the context of the cited paper is not relevant here. That paper states, in the abstract: This study uses Monte Carlo simulations to demonstrate that regression-discontinuity designs arrive at biased estimates *when attributes related to outcomes predict heaping in the running variable*. (emphasis added)

4. The authors use number of beds as importance weights in the hospital-level regressions. If number of beds are important as weights, they may well be important as regression controls. In fact, one would expect hospital size to be significantly related to the number of physicians employed. To be precise, the authors show that there is no evidence of a discontinuity in number of beds across the threshold, suggesting that number of beds as a regression control would not change the results substantively, but that remains an untested proposition. Given the a priori expectation of substantial associations between hospital size and the outcomes, it should be included as a covariate, not just as a weight, in the regressions.

5. The authors use census regions as geographic controls in their primary analyses and states in a supplementary analysis. Both might be considered inadequate. A much sharper quasi experiment would compare hospitals on either side of the DSH threshold in the same market (e.g., hospital referral region (HRR)).

6. The confidence intervals shown in Figure 1 are probably not correct and paint a picture with more statistical precision than is likely true. The authors report 95% confidence intervals, yet the data points (which are group means) are frequently outside the intervals. That strikes me as being incorrect.

Patient level analysis

7. For patient-level analysis, zip code is not a reasonable definition of a market area. Most researchers would use areas defined by Dartmouth (hospital service area (HSA) or HRR). When local policy may have impacts, researchers might use counties or statistical areas as market areas. Zip codes are almost meaningless in any of those contexts.

8. The authors restrict the sample to hospitals that are unique within zip code. They state that this restriction still covers 75% of hospitals. But it means that they drop 25% of hospitals in what is already a restricted sample. Hospitals with nearby competitors might behave quite differently than those without. If that is true, their inclusion could easily change results substantively.

9. Using Medicare claims, I do not understand how one can show impact on low-income patients. Those patients would be enrolled in Medicaid or would be uninsured. This patient level analysis cannot demonstrate whether their care changed in any way. The authors acknowledge this "limitation" in the discussion but it is buried deep at the end of the discussion section.

INTERPRETATION

10. The authors begin their discussion of findings by stating that "... hospitals that are eligible for the 340B Drug Pricing Program have responded to program incentives by ...". This is false, in a strict sense, and more broadly misleading. In fact, their analysis leaves out acute care hospitals with fewer than 50 beds, those with DSH percentages greater than 21.75%, hospitals that are very close to the threshold and several other categories of hospitals like critical access hospitals, sole community hospitals, rural referral centers, pediatric hospitals, and free-standing cancer centers. So, clearly the scope of the interpretation should be limited to that group of hospitals. In fact, to be more precise, because a regression discontinuity design is used, interpretation should be focused on hospitals that are close to the threshold, e.g., "short-term general hospitals close to the threshold have responded to program incentives by ..." To be fair, the authors discuss this and other limitations (limitations paragraphs on p. 9) but these are more than just technical limitations. These speak to policy implications that might be drawn from the work.

11. I found the percentage change interpretations in the paper to be very misleading. For example, going from 1 hematologist-oncologist per hospital to 2.3 hematologist-oncologists per hospital is an increase of 1.3 hematologists-oncologists per hospital. Although it can also be framed as a 230% increase, it clearly makes it feel like it is enormously large and with potentially dire consequences.

Partha Deb, PhD
Professor of Economics
Hunter College and the Graduate Center
CUNY

January 28, 2018