VALUE-ADDED MODELS: NEW METHODS AND APPLICATIONS‡

Sources of Inefficiency in Healthcare and Education†

By Amitabh Chandra and Douglas Staiger*

Healthcare and education share commonalities in mission, financing, and empirical regularities. Both are central to allowing people to perform at their capabilities. Both are often provided by, or subsidized by, government. Heterogeneity—the idea that interventions such as AP classes or bypass surgery may have different effects depending on the characteristics of the recipient and the supplier—is another shared feature of many treatments in education and healthcare. Importantly for this paper, both sectors exhibit wide variation in spending that is loosely associated with outcomes. There are three reasons for this apparent lack of association: (i) confounding—the sorting of students and patients to particular educators or providers in ways that cause nonexperimental measurement of value added to be biased upward or downward; (ii) overuse and underuse—the idea that some schools, hospitals, teachers, or physicians may be doing too much or too little; or (iii) comparative advantage—that some schools or hospitals are better than others in delivering certain types of education or healthcare and so should be delivering more. The last two explanations are related to productivity differences across suppliers, while the first, confounding, is a demand-side phenomenon.

In this paper, we summarize one approach for understanding productivity differences across providers in healthcare and discuss how it can, and cannot, be applied to education. We ignore the role of confounding, which are student-level confounders in education and patient-level confounders in healthcare, not because this is unimportant but because of an appreciation for the considerable progress that has been made in overcoming this challenge (Chetty, Friedman, and Rockoff 2014; Doyle et al. 2016). Rather, we focus on understanding the economic content of variation in risk-adjusted treatment rates or spending. These could arise from two very different mechanisms on the supply side. One interpretation argues that there is a correct amount of use that is the same for everyone, so that variation across providers or educators is evidence of allocative inefficiency—some are using too much care and others are using too little. This would happen if there was overuse at hospitals, possibly because such hospitals were overconfident about the benefit of treatment, or because they were maximizing something other than health. This interpretation of variation leads to an emphasis on guidelines and developing and disseminating information on cost-effectiveness of care. An alternative interpretation is that the return to medical care varies across providers, because of the talents and skills of local providers, so that hospitals with higher returns to treatment should deliver more treatment because of a comparative advantage at doing so. Rather than imposing a uniform standard of care, this interpretation leads to an emphasis on understanding the sources of variation in hospital or physician skill such as support staff and training.

Overuse, underuse, and comparative advantage as explanations for variation in input use have natural counterparts in education. Overuse would occur when a treatment such as AP classes makes the marginal entrant worse off or generates benefits that are less than the social-cost of provision. Underuse would occur if students are unable to access productive teachers or coursework, perhaps because of capacity constraints or imperfect understanding of the benefits of these

‡*Discussants:* Lawrence Katz, Harvard University; Guido Imbens, Stanford University.

*Chandra: Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138 (e-mail: Amitabh_Chandra@Harvard.edu); Staiger: Department of Economics, Rockefeller Hall, Dartmouth College, Hanover, NH 03555 (e-mail: Douglas.Staiger@Dartmouth.edu).
†Go to http://dx.doi.org/10.1257/aer.p20161079 to visit the article page for additional materials and author disclosure statement(s).

interventions. Comparative advantage would suggest that some schools or teachers are better than others in teaching certain types of students, even when facing similar students and with similar resources.

In Chandra and Staiger (2016), we demonstrate that it is possible to separate allocative efficiency (overuse and underuse) from the presence of productive efficiency (comparative advantage) by imposing a prototypical Roy model for the decision to treat a patient. This has some analogues to offering student-specific treatments in education, a point that we return to later. In our model, hospitals choose whether to provide an intensive treatment (reperfusion following a heart attack) based on whether the expected survival benefit for each patient (ΔY_{ih}) exceeds a hospital-specific threshold (τ_h). The expected benefit depends on patient characteristics (X_{ih}), hospital-level expertise at performing treatment (α_h), and an idiosyncratic patient benefit observed by the physician but not the econometrician (ε_{ih}). In our model a patient receives the intensive treatment when $X_{ih}\varphi + \alpha_h + \varepsilon_{ih} > \tau_h$, or equivalently when $X_{ih}\varphi + \theta_h > -\varepsilon_{ih}$ where $\theta_h = \alpha_h - \tau_h$ represents a hospital's risk-adjusted treatment rate (after controlling for patient characteristics in X_{ih}). The minimum threshold (which measures allocative efficiency) should be zero if hospitals only maximize health, but may be negative if hospitals are maximizing something other than health, or are overconfident about the benefits of treatment. The model demonstrates that risk-adjusted treatment rates capture both hospital-specific comparative advantage in the ability to perform treatment and differences in the threshold (allocative efficiency), and are consequently not informative for policy. But it also demonstrates that it is possible to identify allocative inefficiency in the presence of productive efficiency by using an "outcomes test." In particular, it can be shown that the treatment-on-treated estimate for patients receiving reperfusion is given by $E(\Delta Y_{ih}|Treatment_{ih} = 1) = g(I_{ih}) + \tau_h$, where $I_{ih} = X_{ih}\varphi + \theta_h$ is the index determining the propensity to be treated, and $g(\cdot)$ is a tobit-like function that depends on the distribution of ε. Put differently, conditional on a patient's propensity to receive a treatment, any remaining variation in the benefit to treatment across hospitals reflects allocative inefficiency.

Empirically, a patient's propensity to receive treatment is estimated as the prediction from the logit model:

$$(1) \quad \Pr(Treatment_{ih} = 1|X) = F(X_{ih}\varphi + \theta_h),$$

where we estimate the parameters using a mixed logit model, and θ_h is estimated through empirical Bayes. Armed with this propensity, we estimate the presence of allocative efficiency by estimating an outcomes regression for patient i in hospital h:

$$(2) \quad Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \tau_h Treatment_{ih}$$

$$+ \sum_{p=1}^{100} 1(I_{ih} < g_p)Treatment_{ih} + \varepsilon_{ih}^0,$$

where α_h^0 is a hospital fixed-effect and X_{ih} are patient risk adjusters. The benefit from treatment is increasing in the propensity—a fact that is true of a Roy-allocation rule and that we verify. Therefore, we interact the treatment dummy with indicator variables for the 100 percentiles of the propensity to receive treatment. Conditional on these interactions, τ_h is a measure of allocative inefficiency at hospital h. If τ_h is positive at low propensities then it indicates underuse because the marginal patient receives positive benefits. If this term is negative then there is overuse for the marginal patient is harmed as a result of too much treatment. Recovering productive efficiency requires more parametric assumptions which we describe in Chandra and Staiger (2016).

The "single-index" framework from the Roy model is immensely powerful and allows us to combine patient and hospital characteristics into a single index. This is the key theoretical assumption that we need for identification. It may not be trivial for, in words, it assumes that the distribution of the unobservable factors determining treatment is not different across hospitals, which could happen if some hospitals were better at measuring these unobservable factors. Empirically, we need to assume that patients are randomly assigned to hospitals conditional on covariates, but within hospitals doctors triage them according to the benefit that they would receive from treatment. In other words, the random assignment that we require is that treatment is random with respect to baseline unobservables, which may be plausible given

the rich covariates that are at baseline. If there were no allocative inefficiency, then the propensity to receive care for clinically similar patients who are treated at different hospitals would only vary because of differences in a hospital's ability to perform care, but two patients with the same propensity to get care would have the same outcomes (even though these are clinically different patients). But if there is overuse, patients with the same propensity to receive reperfusion treatment will receive lower benefits at hospitals that overuse care (overuse incorrectly increased their propensity for treatment relative to their true propensity, just as underuse incorrectly reduces their propensity).

The graphical intuition for our model can be seen in Figure 1. The expected benefit from treatment is given on the vertical axis, while the propensity of being treated (which depends on I) is given on the horizontal axis; the hurdle is set at zero meaning that providers treat patients until the point of zero-marginal benefit. The top curve in the figure represents the treatment-on-the-treated effect for a patient with a given propensity that is treated in a hospital with a high minimum threshold for treatment ($\tau_{high} > 0$). The lower curve represents the same thing for a hospital with a low minimum threshold ($\tau_{low} > 0$). Treatment-on-the-treated approaches the minimum threshold (τ_{high} or τ_{low}) for a patient with a low propensity of being treated (a very negative I), since no patient is ever treated with a benefit below this threshold. For a patient with a high propensity of being treated (a very positive I), truncation becomes irrelevant and the treatment-on-the-treated effect asymptotes to the unconditional benefit of treatment. However, conditional on a patient's propensity, the treatment effect is always higher in the hospital with the higher threshold.

The graph illustrates three implications of the Roy-selection model. First, we can identify overuse or underuse by focusing on patients with the lowest probability of receiving treatment. In these patients, there is overuse when the treatment effect for the lowest propensity patients is negative, and underuse when the treatment effect for the lowest propensity patients remains positive. Second, differences in hospital specific benefits to performing reperfusion would show up as a movement along the curves—higher comparative advantage at reperfusion increases the propensity of patients to be
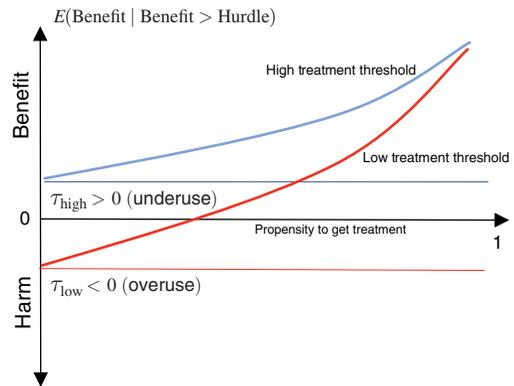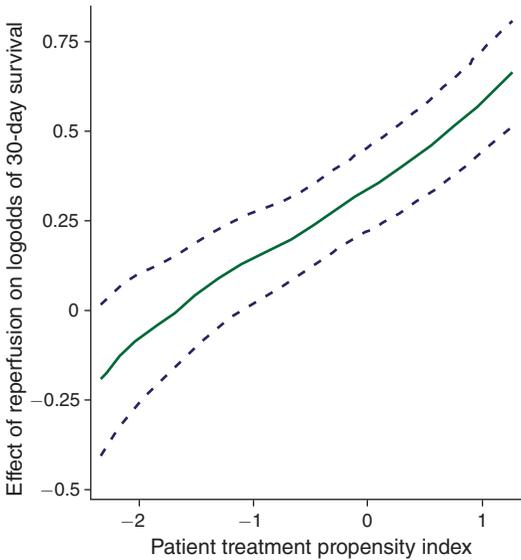

FIGURE 1. IDENTIFYING OVERUSE AND UNDERUSE IN HEALTHCARE

Notes: The curves represent the treatment-on-the-treated effect as a function of the propensity to receive treatment, and approach the minimum threshold (τ) for a patient with a low propensity of being treated. The top curve represents a hospital with a high treatment threshold (underuse) and the bottom curve represents a hospital with a low treatment threshold (overuse).

treated, and therefore the treatment effect, but does not affect treatment effects conditional on propensity. Third, what distinguishes comparative advantage from overuse/underuse is that differences in hospital comparative advantage have an impact on treatment effects by shifting the propensity to be treated, while differences in the minimum threshold have an impact on treatment effects conditional on the propensity to receive reperfusion.

Estimating this model with data for elderly patients following a heart attack, a setting where we have access to chart data and can observe everything that the physician observes, we find strong evidence of allocative inefficiency, with a substantial proportion of hospitals overusing reperfusion therapy among patients who are harmed by the treatment. In Figure 2 we plot the non-parametrically estimated survival benefit from reperfusion against a patient's treatment propensity index for hospitals in the lowest (left-hand side) and highest (right-hand side) terciles of the estimated hospital effect from the propensity equation. Both plots show a strong upward slope, with higher benefit from treatment for patients with a higher propensity to receive reperfusion—and exactly mirrors the theoretical illustration in Figure 1 and validates the Roy

Panel A. Hospitals in lowest tercile
of hospital effects from propensity equation

Panel B. Hospitals in highest tercile
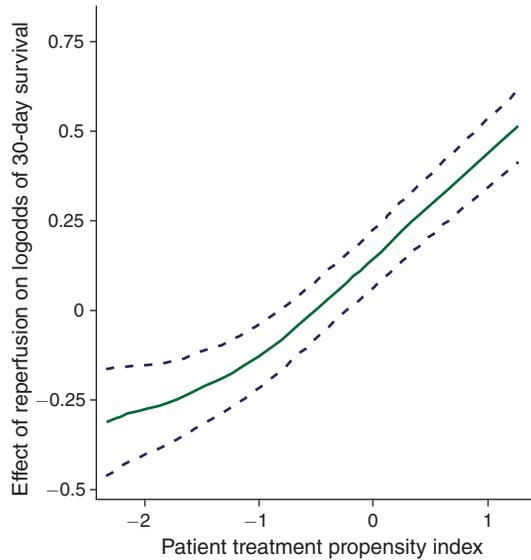of hospital effects from propensity equation

FIGURE 2. SURVIVAL BENEFIT FROM REPERFUSION ACCORDING TO PATIENT'S TREATMENT PROPENSITY

Notes: The figures plot the survival benefit (and 95 percent confidence intervals) from reperfusion against a patient's treatment propensity index for hospitals in the lowest (left-hand side) and highest (right-hand side) terciles of the estimated hospital effect from the propensity equation.

model of treatment assignment. Note that at every propensity, the benefits of reperfusion are lower in the top-tercile hospitals, suggesting that the most aggressive hospitals (with the highest use of treatment) have set a lower threshold for providing care. At the lowest propensity levels, the survival benefits from reperfusion are significantly negative for the top-tercile hospitals, suggesting that there is overuse among these hospitals. In the bottom-tercile hospitals, the estimated survival benefits from reperfusion for the lowest propensity patients are less negative and not significantly different from zero, which is consistent with appropriate use of reperfusion in these hospitals.

In results reported in Chandra and Staiger (2016), we also find substantial variation in hospitals' ability to perform treatment, with the variation across patients and hospitals in the survival benefit from reperfusion being the same order of magnitude as the average treatment effect of reperfusion. We considered two alternative mechanisms that could lead to the allocative inefficiency that we observe in the data. One

possibility is that hospitals are optimizing something other than the survival of a given patient, e.g., over-treating for financial gain (particularly in for-profit hospitals) or because of benefits to future patients through learning-by-doing (particularly in teaching hospitals). This type of mechanism would suggest that allocative inefficiency would be related to hospital characteristics such as ownership, teaching status, etc., but we find no evidence to support this hypothesis. In a second mechanism, allocative inefficiency would arise if hospitals had imperfect information and misperceived their return to treatment. With this mechanism, allocative inefficiency arises because hospitals base treatment decisions on their incorrect perception of the return to treatment in their patients, rather than on the true return to treatment. Given the general lack of systematic performance feedback and small samples of their own treated patients to observe, it is quite plausible that hospitals and physicians will have inaccurate beliefs about their own treatment effectiveness. We find some evidence in favor of this mechanism, with smaller

hospitals having particularly imprecise information about their own treatment effectiveness.

Extending this framework to education is natural for student-level interventions such as AP courses, gifted and talented programs, grade retention, suspension, or use of educational technology. In each of these examples, school = hospital, student = patient, and efficient allocation occurs when students are assigned to treatment (e.g., enrolled in an AP class) based on whether the student's individual benefit from treatment exceeds a common school-level threshold (the opportunity cost). There is lots of variation in these inputs across schools, great interest in estimating treatment effects, and it is reasonable to believe there are very heterogeneous effects across students and schools. In these settings, conditional on the propensity to receive the treatment, evidence of higher returns to, say, AP courses is evidence that a school sets a higher threshold for providing the treatment than other schools, and positive returns among low propensity students would be evidence of underprovision.

The extension to charter schools is less direct, principally because the Roy-model structure of our healthcare model partially breaks down. In this example, school district = hospital, student = patient, and attending charter school or magnet school is the treatment. There are similar motivating facts to healthcare: variation across districts in use of charters, great interest observationally and experimentally in estimating the effect of charters relative to "usual" education in public schools, and belief that there are heterogeneous treatment effects both across students and across districts/charters. The problem is that students are not being allocated in the way our model works, where a student is allotted to charter schools if their individual benefit is over a district-level threshold. Instead they apply to school if their benefit is over a threshold, and then get in by random allocation. Moreover, the threshold in this application would be something like the cost of attending (e.g., based on travel time) compared to the benefit (academic gain), so there is student-level variation in the threshold that researchers could exploit but that also complicates the analysis.

Another example lies in the evaluation of Teach for America (TFA) teachers. In this example, school = hospital, classroom = patient, and TFA teacher is the treatment. Once again, there are similar motivating facts to our healthcare example: variation across schools in use of TFA teachers, great interest in estimating the effect of TFA teachers relative to "usual" teachers, and reasonable to believe that there are heterogeneous treatment effects both across students and across schools. Our Roy framework assumes that principals assign TFA teachers to classrooms that receive the most benefit (e.g., where usual teachers are ineffective). The problem is that schools may not optimally allocate classrooms to teachers based on benefit to the classroom, but instead may simply place TFA teachers in the most difficult to teach classrooms—which regular teachers do not want, but also where unprepared TFA teachers may be the least effective. So the students most likely to get TFA teachers (high treatment propensity students in our framing) may see the most negative treatment effects. Still, this would be an interesting avenue for researchers to explore.

In summary, while research in education has focused on value-added models of teachers and schools, there is still work to be done on understanding what to make of variation in the use of inputs. Here, insights from healthcare may be useful—especially models that use simple economic structure to separate allocative inefficiency from productive inefficiency.

REFERENCES

Chandra, Amitabh, and Douglas O. Staiger 2016. "Allocative and Productive Inefficiency in Healthcare." Unpublished.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Doyle, Joseph, John Graves, Jonathan Gruber, and Samuel Kleiner. 2016. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." Unpublished.