NBER WORKING PAPER SERIES

INTERVENING ON THE DATA TO IMPROVE THE PERFORMANCE OF HEALTH PLAN PAYMENT METHODS

Savannah L. Bergquist
Timothy J. Layton
Thomas G. McGuire
Sherri Rose

Working Paper 24491
http://www.nber.org/papers/w24491

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2018

This research was supported by the National Institute of Aging (P01 AG032952) and the Laura and John Arnold Foundation. Bergquist received additional support from the Harvard Data Science Initiative. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24491.ack

Intervening on the Data to Improve the Performance of Health Plan Payment Methods
Savannah L. Bergquist, Timothy J. Layton, Thomas G. McGuire, and Sherri Rose
NBER Working Paper No. 24491
April 2018
JEL No. I13

ABSTRACT

The conventional method for developing health care plan payment systems uses existing data to study alternative algorithms with the purpose of creating incentives for an efficient and fair health care system. In this paper, we take a different approach and modify the input data rather than the algorithm, so that the data used for calibration reflect the desired levels of spending rather than the observed spending levels typically used for setting health plan payments. We refer to our proposed method as "intervening on the data." We first present a general economic model that incorporates the previously overlooked two-way relationship between health plan payment and insurer actions. We then demonstrate our approach in two applications in Medicare: an inefficiency example focused on underprovision of care for individuals with chronic illnesses, and an unfairness example addressing health care disparities by geographic income levels. We empirically compare intervening on the data to two other methods commonly used to address inefficiencies and disparities: adding risk adjustor variables, and introducing constraints on the risk adjustment coefficients to redirect revenues. Adding risk adjustors, while the most common policy approach, is the least effective method in our applications. Intervening on the data and constrained regression are both effective. The "side effects" of these approaches, though generally positive, vary according to the empirical context. Intervening on the data is an easy-to-use, intuitive approach for addressing economic efficiency and fairness misallocations in individual health insurance markets.

Savannah L. Bergquist
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
bergquist@g.harvard.edu

Timothy J. Layton
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
and NBER
layton@hcp.med.harvard.edu

Thomas G. McGuire
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
and NBER
mcguire@hcp.med.harvard.edu

Sherri Rose
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
rose@hcp.med.harvard.edu

1. Introduction

Public and private regulators set prices in health care – to physicians, hospitals, health plans and other providers – with the purpose of improving economic efficiency of health care delivery. To implement a payment system, the regulator selects a methodology to establish prices for each of the thousands of procedures physicians perform, the hundreds of disease groups treated by hospitals, or per capita prices paid to plans for the nearly innumerable combinations of enrollees' health conditions. In many health care settings, prices are set according to observed costs in the market. In the U.S. Medicare program, a hospital's payment for a given inpatient admission is tied to the national average cost of admissions belonging to the same diagnosis-related group (DRG). In the Medicare Advantage program, the Affordable Care Act Marketplaces, most state Medicaid programs, and in Germany, the Netherlands, Switzerland, and other countries, a health plan's payment for enrolling a given individual is tied to the average cost of individuals with similar demographics and histories of chronic conditions. This method for setting prices weakens incentives for inefficient "cream-skimming" behaviors by insurers (Geruso, Layton and Prinz 2016; Newhouse, et al. 2015).

While acknowledging the benefits of setting prices in this way, this paper explores a previously overlooked deficiency: when the existing health care system is inefficient or unfair, setting payments based on observed costs may sustain those inefficiencies or inequities, which may be precisely the issues the payment system is meant to correct. Consider the following extreme example: suppose under the current health plan payment system individuals with cancer are highly unprofitable due to low revenues relative to costs, and insurers rationally respond by limiting access to care for individuals with cancer to deter those individuals from enrolling in their plans. Further, assume insurers are so effective at rationing access that individuals with cancer obtain only very low levels of spending, for example, $1,000. Now suppose a regulator seeks to address the problem of insurers inefficiently limiting access to care for individuals with cancer by "risk adjusting" payments to health plans, and they use a standard risk adjustment system where plan payments are adjusted according to the observed $1,000 of spending for individuals with cancer. Such a payment system will not solve the problem the regulator sought to fix, because individuals with cancer will appear relatively inexpensive due to the insurers' actions limiting their access to care. Thus, this conventional approach to payment will sustain rather than correct the insurers' incentive to inefficiently limit access to care for this group. While this example is extreme, a weaker version of this feedback loop between inefficiencies embedded in the health care system and the incentives embedded in the payments is likely to play out in many more realistic settings.[1] The general point is that if regulated prices are intended to move the health care system to be more efficient and fair, using existing (inefficient/unfair) patterns of care for purposes of payment calibration is unlikely to be the right approach.

In this paper, we show that data modified to reflect the researcher's or policymaker's beliefs about *efficient* and *fair* levels of spending versus the *observed* spending levels can be used for calibrating health plan payments. We refer to this approach as "intervening on the data." The intuition behind intervening on the data follows from conditions of market equilibrium: set prices to buy the equilibrium you want, not the equilibrium you currently have. After introducing background material in Section 2, we lay out the theoretical basis of our approach in Section 3. The model in Section 3 characterizes the optimal input data (that which leads to the desired health care system outcomes) at an abstract level, and then translates the ideas for our empirical applications of health plan payment in Medicare. In Section 4, we describe two examples of intervening on the data. The first focuses on underpayment for individuals with chronic illnesses, and the second addresses disparities in health care spending by zip code-level income. We empirically compare intervening on the data to approaches based on modifying payment algorithms using existing data. This includes exploring the conventional approach of

[1] Algorithmic fairness with biased data is a growing area of study outside of health care. The theme of using a data- or algorithm-based intervention to break a feedback loop that reproduces or exacerbates biases in data runs through many, diverse applications. For example, Lum and Isaac (2016) discuss how using biased data in predictive policing algorithms can reproduce and amplify the biases in police data, and in some cases lead to a discriminatory feedback loop between policing practices and algorithm predictions. Hu and Chen (2017) show how a fairness constraint can be imposed on firms' hiring practices to ensure both individual- and group-based fairness and address discrimination in the labor market.

adding or removing risk adjustor variables from the pricing algorithm. In the chronic diseases example, we include additional risk adjustors for chronic illness groups, and in the disparities example we include a low-income neighborhood indicator. We also leave the set of risk adjustors in place but introduce constraints on the OLS regression that target spending for groups of interest (van Kleef, et al. 2017).

Both applications demonstrate that intervening on the data is a powerful approach for moving plan payments closer to the desired levels. In the chronic diseases application, intervening on the data is the most effective method for improving fit across all groups. Introducing constraints on the risk adjustment formula also performs well at the group-fit level, while adding risk adjustors to the OLS regression is the least effective approach. The substantial improvements in group-level fit for both intervening on the data and constrained regression come at only a trivial decrease in statistical fit at the individual level (i.e., R^2) compared to an unconstrained OLS regression. In the disparities application, adding a risk adjustor for neighborhood income is highly ineffective if the objective is to address health care spending disparities for that group. Such an action aggravates the disparity by income group. Intervening on the data improves incentives to serve a group, but combining intervening on the data with adding a risk adjustor is even more effective at altering incentives because it combines the strengths of both approaches.

2. Background

Medicare Health Plan Payment and Risk Adjustment

Medicare adjusts per person payments to health plans according to age, sex, Medicaid status, reason for Medicare eligibility, and diagnostic information. The Medicare risk adjustment formula is calibrated on data from traditional fee-for-service Medicare but used to pay private Medicare Advantage plans (whose data are not used for calibration purposes). Most efforts to improve risk adjustment have centered on adding and refining the risk adjustor variables that indicate the presence of a diagnosis for a particular chronic disease (currently the Hierarchical Condition Categories, abbreviated HCCs), or on segmenting the population and estimating mutually exclusive subgroup formulas (e.g., separating community-dwelling and institutionalized beneficiaries). Medicare tracks regression fit at the individual level, updating the risk adjustment coefficients every few years by refitting the regression on more recent fee-for-service data. The most recently publicly evaluated Medicare risk adjustment formula is CMS-HCC Version 21; the subgroup formula for the largest group, aged beneficiaries, has an R^2 of 12.46%.[2]

Despite a risk adjustment scheme refined over many years, some groups remain vulnerable to selection-related incentives. For example, plans are underpaid for individuals with certain chronic conditions in that the revenue a plan receives for members of this group falls short of the expected costs. Research on Medicare Advantage plans shows plans react to such "underpayments" by cutting back on services to unprofitable illness groups (Han and Lavetti 2017).

Intervening on the Data as a Policy Tool

The most significant example in the U.S. of intervening on the data for setting provider payment is the Resource-Based Relative Value Scale (RBRVS) used by Medicare since the early 1990s to set physician procedure prices. Prior to the RBRVS reform, a policy consensus held that the prices set for "Evaluation and Management" (E&M) procedures (i.e., routine office visits) were too low and prices for some interventions (i.e., many surgical procedures) were too high. These payment levels resulted in too little physician time spent with patients and too high a volume of some procedures – patterns in the existing data policymakers wanted to change (Newhouse 2002). Rather than accepting these patterns, Medicare incorporated information about normative time and effort – what doctors should be doing – and raised the prices of E&M

[2] The Medicare risk adjustment model currently used in practice is CMS-HCC Version 22; it includes fewer HCCs than Version 21 and estimates a greater number of subgroup models.

procedures relative to surgery. In short, the RBRVS fee schedule was designed not to simply *reflect* the time spent by physicians for various procedures, it was meant to *affect* the time spent.[3]

We also see limited forms of intervening on the data in regulated health insurance markets in the Netherlands and the ACA Marketplaces. The Dutch risk adjustment algorithm is estimated on data modified to incorporate benefit package changes over time (Layton, McGuire and van Kleef 2016). In the ACA Marketplaces, the risk adjustment algorithm is estimated on a pooled sample of data not subject to the selection incentives in the Marketplaces, and modified by an anticipated annual growth rate. The 2017 update of the formula additionally features separate specialty drug, traditional drug, and medical expenditure trends based on external actuary and industry reports. Similar to the Dutch approach, the goal of applying a trend modification to the data is to incorporate expected patterns of spending, not to alter those expected patterns (Centers for Medicare and Medicaid Services 2016).

3. Conceptual Framework

Here, we develop a general model of plan actions in response to plan payment coupled with recognition that plan payments are endogenous to plan actions. We explain our idea of intervening on the data, and then describe how it can be used for addressing spending allocation problems.

A Model of Intervening on the Data

Define the matrix $\mathbf{X} = \{x_{is}\}$, where *i* indexes individuals in the health insurance market and *s* indexes the services provided by health plans. A service could be a particular type of care for a particular condition or conditions, e.g., office-based care for depression. An element of $\mathbf{X}$ is x_{is},

[3] This is an example of the well-accepted idea that, as Newhouse (2002) pointed out in the context of physician pricing, provider costs are endogenous to the payment method.

spending in dollars on service *s* for individual *i*. Also define $\mathbf{R} = \{r_i\}$, a vector of payments a health plan receives for enrolling each individual *i*. $\mathbf{X}$ is the performance of the health care system and $\mathbf{R}$ is the set of payments chosen by the regulator to influence the system.

$\mathbf{X}$ and $\mathbf{R}$ are related in two ways. The vector of payments affects the health care system. And data from the health care system are inputs into the algorithm that generates the payment vector. We represent these two relations as:

$$\mathbf{X} = \mathrm{f}(\mathbf{R}) \qquad (1)$$

$$\mathbf{R} = \mathrm{g}(\mathbf{X}) \qquad (2)$$

The function f(·) represents the working of the health care system in response to payment incentives. Dependent on the payments set for each enrollee, plans set contracts with providers and engage in other actions, including selection, that lead to an allocation of spending on various services to each individual. The function g(·) is the algorithm specified by the regulator that takes as input the patterns of health care spending and determines a payment for each person. The empirical methodology used by Medicare to assign risk scores to individuals based on age and indictors of service use is an example of a g(·) function.[4]

In the conventional approach to economic analysis of health plan payment, the payment vector $\mathbf{R}$ is treated as an exogenous policy choice, and the health care system is characterized by f(·) only. This may not, however, constitute a full equilibrium. In general, the health care system that results from a given payment vector will not, given the payment algorithm g(·), generate the original payment vector. We broaden the concept of equilibrium by recognizing the second relationship between the health care system and payments, represented in g(·). We say that with a fixed algorithm g(·), relating data from the health care system to payment, the system is

[4] Relations (1) and (2) should be understood broadly and remain true even in a system like Medicare where the payment system calibration is based on data from fee-for-service Medicare. Medicare Advantage plan behavior in response to payment incentives affects how many and what type of beneficiaries remain in fee-for-service Medicare. The relations also remain in place even in the case in which the data used for calibration is entirely disconnected from system performance. In that case, the g(**X**) function is a set of payments unaffected by **X**.

in full equilibrium when both conditions are satisfied. We thus define a full equilibrium in the health care system $(\mathbf{X^e}, \mathbf{R^e})$ as:[5]

$$\mathbf{X^e} = \mathrm{f}(\mathbf{R^e}) \qquad (3)$$

$$\mathbf{R^e} = \mathrm{g}(\mathbf{X^e}) \qquad (4)$$

To motivate intervening on the data, we assume another allocation, $\mathbf{X}^*$, is preferred to $\mathbf{X^e}$. We refer to $\mathbf{X}^*$ as the "desired" allocation; $\mathbf{X}^*$might be preferred to $\mathbf{X^e}$ for reasons of efficiency or fairness.

Conventional health policy analysis modifies the g(·) function (for example, by adding variables to the risk adjustment regression). With knowledge of the operation of the health care system (i.e., the f(·) function), one could solve for the vector of per person payments that would lead to the desired allocation.[6] Call this desired vector of payments $\mathbf{R}^*$:

$$\mathbf{R}^* = \mathrm{f}^{-1}(\mathbf{X}^*) \qquad (5)$$

Equation (5) represents the strand of the health plan payment literature referred to as "optimal risk adjustment," with f^{-1} as the optimal risk adjustment formula.[7] Note that (5) is consistent with a full equilibrium, in the sense that the optimal risk adjustment formula, f^{-1}, when applied to the optimal health care system, $\mathbf{X}^*$, yields the optimal vector of per person payments $\mathbf{R}^*$.

In this work, we take a different approach and ask, what $\mathbf{X}$ matrix, used in equation (2) with a fixed algorithm g(·), would lead to the optimal set of per person payments, and thus to the efficient health care system. Define $\ddot{\mathbf{X}}$ to be this optimal data input matrix:

---

[5] In practice, data from a previous period are typically used to set payments in the current period. If a system is in full equilibrium as we describe, such timing does not affect the equilibrium. Out of equilibrium, a system characterized by (1) and (2) would need to take account of timing and any iterative process leading to equilibrium.

[6] Throughout, we assume that solutions exist, correspond to positive values of all spending, and are unique and stable. Our Medicare application only requires the assumptions to hold locally.

[7] In Glazer and McGuire (2002), optimal coefficients are those that lead the plan, in profit maximization, to supply efficient health care. In both of these instances, the coefficients are found by inverting the function describing system response to the price incentives. Theoretical solutions to optimal risk adjustment are assumed to be applied to the optimal allocations and are not designed to move the system to optimal based upon some arbitrary inefficient/unfair initial allocation.

$\ddot{\mathbf{X}} = \mathrm{g}^{-1}(\mathbf{R}^*)$, or

$\ddot{\mathbf{X}} = \mathrm{g}^{-1}\left(\mathrm{f}^{-1}(\mathbf{X}^*)\right)$, or

$$\ddot{\mathbf{X}} = \mathrm{h}(\mathbf{X}^*) \quad (6)$$

Equation (6) presents our main idea in abstract form. By disconnecting the data used as input to the plan payment algorithm from the spending in the current system, i.e., *by intervening on the data,* we can induce an efficient health care system. We recognize the h(·) function is complex and we have little policy experience with its properties. Nonetheless, we believe (6) represents a novel and useful observation. Data used as input into payment system calibration, $\ddot{\mathbf{X}}$ in (6), can be chosen by policy. And although the form of (6) is unknown, simple and intuitive modifications to the input data have foreseeable effects on incentives. The correct intervention on the data that provides insurers with the right set of incentives will at least move the health care system in the right direction.[8]

A Medicare Application

To address allocation problems in Medicare we operationalize (6) and illustrate intervening on the data. We specify the variables and some properties of how the health care system responds to price incentives (f(·)), and how the data from the health care system is used to construct payments to health plans (g(·)). In this section, we use the chronic diseases application to translate our model into practice. A parallel method is used in our disparities application. [9]

[8] There are direct parallels in our proposed approach of intervening on the data to the theory underlying structural causal modeling (Pearl 2009). Specifically, we could formulate the two relationships in (1) and (2) as nonparametric structural equation models with unknown functional forms f(·) and g(·), although an explicit representation of equilibrium does not always apply. We can intervene on the system we have defined to set random variables (here, payments) to specific values.

[9] We focus on a subset of features of Medicare payment and Medicare Advantage plan spending allocations. Our examples are concerned only with medical spending predicted by the Medicare Advantage risk adjustment formula – primary Part A and B spending – not drug spending (which has its own plan payment system). The analyses we present focus on community-dwelling individuals, the largest set of Medicare beneficiaries. We ignore any joint cost issues introduced by the presence of other payers of health plan services.

The Inefficiency

Based on previous research, we start with the premise that the Medicare payment system to health plans creates incentives to underspend on persons with some chronic illnesses. As McGuire et al. (2014) show in another context and our data analysis confirms for Medicare, among the four chronic illness groups we study – cancer, diabetes, heart disease and mental illness – the incentives are strongest to underspend on mental health care. We therefore intervene to address the incentives to underspend on persons with mental illness, and track the incentives for groups defined by the other chronic illnesses.

The Plan Payment Algorithm: $\mathbf{R} = \mathrm{g}(\mathbf{X})$

The function g(·) is the algorithm that determines the per person payments paid to health plans for enrollment of beneficiaries with various indicators for health status (which come from the data). We implement a slightly modified version of Medicare's current algorithm, described in detail in Section 4 and the Appendix.

The Health Care System: $\mathbf{X} = \mathrm{f}(\mathbf{R})$

The $\mathrm{f}(\mathbf{R})$ function relates per person payments to the functioning of the health care system. This function encompasses not only plan spending and beneficiary enrollment, but also involves the relationships between providers, hospitals, insurers, and individual consumers. Rather than making strong assumptions about the form of f(·), we make a weaker assumption about its local properties, the signs of the derivatives of f(·). This assumption is sufficient to determine the sign of the local effect of a data intervention. Specifically, we assume plans respond positively to incentives in the sense that if the payment for an individual is increased (decreased), the plan spends more (less) on that person.[10] The assumption implies that in

[10] This assumption is based on the model of health plan choice and plan design presented in Frank, Glazer, and McGuire (2000). Geruso, Layton, and Prinz (2016), Carey (2017), Lavetti and Simon (Forthcoming), and Shepard (2016) present empirical evidence supporting this assumption.

response to an increase in the level of plan payment for a group of beneficiaries, plans will spend more on the services used by those beneficiaries in order to encourage more of that group to enroll.

4. Data and Methods

We study methods for reducing underpayment (which we assume induces underservice) for the targeted group in the chronic diseases application, and methods to reduce underservice for the targeted group in the disparities application. In both the chronic diseases and disparities applications, we identify the target group using a classification system that is distinct from the CMS-HCC risk adjustment system. We track the impact of alternative methods on regression fit at the group and individual levels. As a summary measure, we calculate group payment system fit (GPSF), which captures how well a given risk adjustment formula matches payments and costs at the group, rather than the individual, level (Layton, et al. 2017).

Data and Baseline Risk Adjustment Formula

In practice and in this paper, data from fee-for-service Medicare are used to estimate payment coefficients on the risk adjustment variables for setting capitation payments to Medicare Advantage plans. We follow CMS and include only individuals with 12 months of continuous Parts A and B coverage in the base year (2010) and those with at least one month of coverage in the prediction year (2011). We focus on beneficiaries who are entitled to Medicare on the basis of age, and exclude those with end-stage renal disease, and those residing in an institution. After applying these exclusion criteria, we draw a random sample of 1.5 million individuals in order to approximate CMS's practice of using a random 5% sample of Medicare beneficiaries for risk adjustment estimation. Our sample is predominately white (85%) and urban-dwelling (76%) (Table 1). Approximately two-thirds of our sample has at least one of four major chronic conditions: cancer, diabetes, heart disease, or mental illness. We combine the 20% random sample Medicare Beneficiary Summary File on enrollment with Part A, Part B, home health, and durable medical equipment claims files for 2010 and 2011. We map ICD-9 diagnoses to 87 HCCs, combine age and sex into 24 age-sex cells, and create indicators for disabled enrollees and individuals enrolled in Medicaid. Our outcome measure, health care spending in 2011, is total Medicare spending on inpatient, outpatient, durable medical equipment, and home health, weighted to reflect each individual's eligible fraction of the prediction year. To predict spending we use a similar specification as the Version 21 CMS-HCC risk adjustment formula for the aged, community dwelling subgroup.

We map ICD-9 diagnoses to Clinical Classification Software (CCS) groups and to Prescription Drug Hierarchical Condition Categories (RxHCCs) (AHRQ 2016; DHHS 2017). CCS groups are used to identify beneficiaries with a chronic condition because, unlike the HCCs included in the risk adjustment formula, all ICD-9 codes map to a condition group.[11] RxHCCs are the categories used by the Medicare Part D risk adjustment formula, and capture a greater number of mental illness diagnoses than HCCs (Montz, et al. 2016). We use the RxHCCs to demonstrate the effect of inclusion of additional risk adjustor variables.

Defining and Measuring Underpayment for a Group with a Chronic Illness

In our first application we target the systematic underpayment of persons with a chronic illness. We divide our sample into seven mutually exclusive groups based on clinical diagnoses in CCS categories: 1) mental illness only, 2) mental illness and at least one other major chronic condition (diabetes, heart disease, or cancer), 3) diabetes only, 4) heart disease only, 5) cancer only, 6) multiple chronic conditions, not including mental illness, and 7) no chronic conditions.

[11] In our sample, the mental illness HCCs capture only 44% of individuals in the mental illness group defined by CCS groups. The difference in how CCS groups are defined versus how the HCC-based risk adjustors are defined accounts for why some groups are underpaid in the current risk adjustment scheme, despite having HCC indicators in the formula.

We define net compensation as the difference between average payments and average costs for a group.[12] We assume an individual's total Medicare spending is the cost the plan bears for that individual, and payments to a plan are the predicted values from the risk adjustment regression: $r_i = \sum_i \hat{\beta}_j z_{i,j}$, where r_i is the payment for individual *i*, $\hat{\beta}_j$ is the estimated coefficient for risk adjuster *j*, and z_{ij} is the value of the risk adjuster *j* for individual *i*. We then calculate group-level net compensation for the average plan based on the predicted values:

$$\text{Net compensation}_g = r_g - c_g, \qquad (7)$$

where $r_g = \sum_{i \in g} r_i / n_g$ and $c_g = \sum_{i \in g} Y_i / n_g$. Negative values of net compensation indicate underpayment, where the group-level revenues (predicted spending) are systematically less than group-level costs (observed spending). Positive values indicate overpayment, where group-level revenues exceed group-level costs.[13]

Defining and Measuring a Health Care Disparity

Our second application addresses a disparity between groups defined by a sociodemographic variable. We divide our sample into individuals residing in low-income zip codes (those with a median income in the lower 60% of our sample zip code median incomes) and high-income zip codes (upper 40%).[14] We measure the disparity in health care between these two groups. Measuring and addressing disparities between groups does not require knowledge about the source(s) of the disparity – many factors contribute to disparities in care across neighborhood of residence. Rather than seeking to directly address these underlying causes, we rely on the

[12] Papers and government evaluations of underpayment by group in the U.S. commonly measure underpayment in ratio form (predictive ratios) rather than by a difference (e.g., McGuire, et al. 2014; Pope, et al. 2011). Here we work with the difference metric to keep net compensation on the dollar scale. Our approach is more in line with the European literature, which uses the difference rather than the ratio (e.g., van Kleef, van Vliet and van de Ven 2013).

[13] Note that net compensation is defined in relation to *observed* spending rather than the desired, efficient level of spending. We return to the matter of using observed vs. efficient levels for payment system evaluation in the Discussion.

[14] We examined disparities across zip code median income quintiles, and compared two different definitions for the reference group: the upper 60% and the upper 40%. We found the middle quintile aligned more closely with the bottom two quintiles, thus we defined the reference group (high-income) as the upper 40% and the group of interest (low-income) as the lower 60%.

responsiveness of the health care system to financial incentives to reduce disparities between low- and high-income neighborhoods: if health plans are paid more to enroll individuals in low-income neighborhoods, then the plans will act to attract these individuals by providing them with more care.

Disparity for a group is defined, following the Institute of Medicine's *Unequal Treatment* Report (2002), as differences in health care not due to health status or preferences. Because of the difficulties in measuring preferences, the literature on disparities compares care for groups after adjusting for health status only,[15] and we do the same. Specifically, we estimate a disparity in total health care spending between the group of interest and the reference group as the difference in estimated and observed price-adjusted spending, conditional on health status. We define disparity as:

$$D_{g,ref} = \frac{1}{n_g}\sum_{i\in g}(Y_{i,g} - \hat{Y}_{i,g}) - \frac{1}{n_{ref}}\sum_{i\in ref}\left(Y_{i,ref} - \hat{Y}_{i,ref}\right), \quad (8)$$

where *g* is the group of interest, *ref* is the reference group, $Y_{i,g}$ is the observed spending for an individual in the group of interest, $Y_{i,ref}$ is the observed spending for an individual in the reference group, $\hat{Y}_{i,g}$ is predicted spending conditional on health status (as measured by CCS categories, age, and sex) for an individual in group *g*, and $\hat{Y}_{i,ref}$ is predicted spending conditional on health status for an individual in the reference group. We follow common practice in the disparities literature and also adjust for age and sex, because differences related to these demographic factors could stem from differences in prevalence of illness correlated with age and sex, and thus would not constitute a disparity. The first term in (8), for group *g* (low-income), is expected to be negative, with actual services received less than predicted based on health status. The second term is expected to be positive, measuring the excess of services received by the reference group in relation to their health status. When the difference between these two terms, the disparity, is negative, it measures the underservice of group *g,* conditional on health status, in relation to the reference group.

[15] For a review of methods for implementing the IOM definition of health care disparity, see Cook, McGuire, and Zaslavksy (2012).

The estimation of $D_{g,ref}$ in equation (8) depends on accurate predicted values $\widehat{Y_i}$, where "accuracy" can be defined with respect to an underlying squared error loss function.[16] Therefore, we employ ensemble statistical machine learning methods to obtain optimal predicted values in the calculation of the disparity. The Appendix contains additional details on implementation. As far as we know, this is the first application of big-data methods to estimating health care disparities.

Intervening on the Data

<u>Targeting Underpayment for a Group with a Chronic Illness</u>

In our demonstration, first intervene on the mental illness group, as identified by CCS category. Specifically, starting with the basic OLS specification, $Y_i = \sum_i \beta_j z_{ij}$, we increase the *Y* vector by 10% for each individual in the targeted group. The choice of 10% is arbitrary. Notably, we only need to intervene on the data once in order to project the effect of changing the transfer by any amount because of the linear form of least-squares estimators.[17] This linearity property does not, however, hold for the impact of data changes on a quadratic measure of fit.

After intervening to transform the *Y* vector, we reestimate the regression, ensuring total payments remain constant by imposing a constraint on the coefficients:

$$\bar{Y}_N = \sum_j \beta_j \, \bar{z}_{j,N}, \qquad (9)$$

[16] Other loss functions can be considered, including those for bounded continuous outcomes.

[17] Intervening on Y_i leads to a constant change in net compensation because revenues at the individual and group level are linear in *Y*:

$r_g = \bar{\hat{Y}}_g$

$\bar{\hat{Y}}_g = \hat{\beta} Z_g$

$\hat{\beta} = (Z'Z)^{-1} Z'Y$

Intervening on the data changes *Y*, while $(Z'Z)^{-1}Z'$ and c_g are constant. Thus, any ΔY leads to the same $\Delta\beta$, $\Delta\hat{Y}$, and, ultimately, change in net compensation.

where $\bar{Y}_N$ is the sample average of observed total Medicare spending and $\bar{z}_{j,N}$ are the sample means for the *j* risk adjustor indicator variables. By constraining the coefficients to produce the observed sample mean spending, we guarantee that the total payments after the data intervention sum to total costs. Put differently, we implement the constraint (9) so that spending for individuals without a mental illness is reduced and counterbalances the increase for individuals with a mental illness.[18] From the refitted regression we obtain a new set of predicted revenues for each individual in the sample.

Using these new predicted spending values, we can then recalculate net compensation for each disease subgroup in the manner described above. Net compensation reveals how a given group fares under a particular risk adjustment formula, but does not summarize overall group fit across the entire sample. We use group payment system fit (GPSF) (Layton, et al. 2017) to summarize the overall impact of each risk adjustment approach on net compensation:

$$\text{GPSF} = 1 - \frac{\sum_g s_g |c_g - r_g|}{\sum_g s_g |c_g - \bar{c}|} \qquad (10)$$

where r_g comes from the estimated risk adjustment regression and c_g is the observed, unmodified vector of spending; $\bar{c}$ is the mean of the observed, unmodified vector of spending for the entire sample; and s_g is the share of the sample in group g, with $\sum_g s_g = 1$. GPSF is a measure of regression fit calculated at the group-level; we use a linear scale to maintain consistency with our net compensation measure.[19] The denominator is the weighted sum of absolute differences between group-level costs and the overall sample average, and the numerator is the group-level sum of absolute residuals under a given payment system. GPSF falls between 0 (the risk adjustment formula is equivalent to paying sample average costs) and 1 (the risk adjustment formula perfectly matches payments and costs at the group level). We

[18] Another way to maintain the same aggregate budget would be to directly deduct the corresponding amount of spending from the complementary group. Using a constraint, however, is our preferred approach. First, using a constraint does not require calculating the addition to the target group before running the regression, allowing the analysis to proceed in one step rather than two. And second, if the intervention on the data is complex – for example, if there are multiple targeted groups – our approach of intervening on the *Y* vector to enact the desired changes for the targeted groups and then imposing an overall constraint is simple and easily implemented.

[19] GPSF on a linear scale is a group-level version of Cumming's Prediction Measure (CPM) (van Veen, et al. 2015).

expect the denominator to be large relative to the numerator (and GPSF to be high) in cases where group-level costs vary widely from the average cost of the entire sample, and group-level payments are close to group-level costs. In our application, GPSF calculated using the baseline risk adjustment formula is the floor we seek to improve on with alternative risk adjustment approaches, and 1 is the ceiling. Comparing GPSF allows us to rank our risk adjustment approaches based on how well they address discrepancies in net compensation for all of the defined groups, not just the targeted group of interest. Concerns about how other groups might be affected by intervening on the data or imposing linear constraints should guide the selection of the mutually exclusive groups for calculating GPSF.

To compare fit at the individual level of the reestimated regression to the baseline OLS version, we calculate an R^2 as follows:

$$\mathrm{SSR} = \sum\left(Y_{i,obs} - \hat{Y}_{i,refit}\right)^2,$$

$$\mathrm{SST} = \sum\left(\mathrm{Y}_{\mathrm{i},obs} - \overline{\mathrm{Y}}_{obs}\right)^2,$$

$$\mathrm{R}^2 = 1 - \mathrm{SSR}/\mathrm{SST}, \qquad (11)$$

where $\hat{Y}_{i,refit}$ comes from the refit regression; $Y_{i,obs}$ is the observed, unmodified vector of spending; and $\overline{Y}_{obs}$ is the mean of $Y_{i,obs}$.

In addition to specifically targeting one chronic illness group, we perform another intervention, where we proportionally increase spending for a beneficiary with at least one of the chronic conditions. Here the goal is to provide an additional example to demonstrate how intervening on the data based on a group categorization can affect subgroups within the broader classification. We estimate the effect of an initial 5% increase, and show proportional increases in the range of 0-15% of spending.

Targeting a Disparity

In the disparities application, we use the plan payment system to reduce incentives for the targeted underservice. Recall that the method is agnostic as to the cause of the disparity in the first place, but makes the simple assumption, discussed above, that increasing the profitability of the group suffering a disparity will encourage plans to provide that group with more services. Because we do not modify costs, tracking changes in group-level payments is sufficient to capture how well intervening on the data addresses disparities. We calculate group-level payments, $r_g = \sum_{i \in g} r_i / n_g$, by estimating the same CMS-HCC risk adjustment formula described above to obtain individual level payments (r_i). Groups that are underserved, like those living in low-income areas, will likely already have payments exceed costs, i.e., be profitable. Despite this, the disparity indicates that the incentives conveyed by the payment system are not strong enough to encourage health plans to compete for these individuals by providing them with more services, and thus eliminate disparities.

To intervene on the data, we choose to increase spending by 10% for all individuals in the low-income zip code group, and again reduce spending for the complementary group by imposing an overall budget constraint. We reestimate the risk adjustment regression with the modified *Y* vector and recalculate group-level payments. It is then possible to characterize the tradeoffs between reducing incentives for disparities and traditional measures of statistical fit for the risk adjustment system, such as R^2, described above in (11). We include an extension of this example in the Appendix: in an even more targeted intervention, we transfer 10% to the low-income group, but take funds only from the upper half of the high-income group (the fifth quintile).

Comparison to Alternative Methods: Adding Risk Adjustor Variables and Constrained Regression

We compare intervening on the data with two changes to the risk algorithm used to define payments based on the existing data. First, in the mental illness application, we add RxHCCs from the Medicare Part D risk adjustment formula related to mental illness and not already included in the Medicare Advantage risk adjustment formula.[20] In the disparities application, we add a low-income neighborhood indicator to the formula. While this latter change is a poor choice because it reinforces incentives to underserve, we do so to illustrate that adding indicators related to socioeconomic status can often be counterproductive and to allow us to set up a combination approach.

Second, we impose linear constraints on the estimated coefficients from the risk adjustment regression to ensure group-level payments by chronic illness group hit targeted levels. Thus, to completely eliminate underpayment for the targeted chronic illness group we set $\bar{Y}_g = \sum_j \beta_j \, \bar{z}_{g,j}$. To compare constrained regression to intervening on the data, we also apply a constraint with the same reduction of underpayment as achieved by intervening on the data. Again, we maintain aggregate spending levels by imposing a second constraint on spending for the entire sample: $\bar{Y}_N = \sum_j \beta_j \, \bar{z}_{N,j}$. We do not implement a constrained regression comparator for the disparities application because it is not possible to directly constrain the target (disparities) without making strong assumptions regarding how the health care system reacts to changes in payments.

5. Results

Increasing Net Compensation for Beneficiaries with a Chronic Illness

Prior to intervening on the data, we assess net compensation by chronic illness. In aggregate, net compensation for individuals with a mental illness is -$652 per person; separating this group into those with mental illness only and those with mental illness and another chronic condition shows the latter and larger group is subject to the greatest underpayment (Table 2). Individuals with heart disease only are the most accurately compensated of the chronic illness groups we examine, followed closely by the multiple chronic conditions with no mental illness

[20] Detail on the selection of the additional risk adjustors can be found in the Appendix.

group. Those with diabetes only, cancer only, and no chronic conditions are overpaid at $768, $247, and $294 per person.

After intervening on the data by increasing spending by 10% for all individuals with a mental illness and refitting the risk adjustment regression, net compensation for individuals with multiple chronic conditions and mental illness falls from -$1,502 to -$291 per person (Table 3, Columns (1) and (2)). Net compensation for individuals with only a mental illness increases from $315 to $644 per person. Notably, compensation for individuals with multiple chronic conditions but no mental illness, cancer only, and no chronic conditions becomes much more accurate. Net compensation for the diabetes only group is reduced, and underpayment is aggravated slightly for the heart disease only group.

Comparison with Alternative Methods Increasing Net Compensation for Beneficiaries with a Chronic Illness

After adding three RxHCC risk adjustors to the baseline regression, average per person net compensation for the multiple chronic conditions with a mental illness group is reduced to -$825. In the mental illness only group, net compensation is increased to $518 per person (Table 3, Column (4)).

Setting the constrained regression target to be equal to the average spending for the mental illness group achieved by intervening on the data (so that the two methods are directly comparable) does more than intervening on the data to increase net compensation for the mental illness only group ($644 vs. $988; Table 3, Columns (2) and (5)) and less to decrease net compensation for the multiple chronic conditions with a mental illness group (-$291 vs. -$433). The constrained regression targeting zero net compensation for all individuals with a mental illness extends these trends, yielding net compensation of $1,006 per person in the mental illness only group, and -$416 per person in the multiple chronic condition with a mental illness group (Table 3, Column (6)).

To calculate GPSF we combine the "mental illness only" and "multiple chronic conditions with a mental illness" group into a single category, which reflects the group targeted by the intervention. At baseline GPSF is 92.032, leaving a gap of 7.968 between the baseline and maximum GPSF. Intervening on the data with a 10% spending transfer yields the largest improvement over the baseline (by 6 percentage points) with a GPSF of 98.009. The constrained regression approach leads to a 5 percentage point improvement over baseline, while adding in the risk adjustors generates only a 1 percentage point improvement. In contrast, adding the risk adjustors has the best individual fit, with an R^2 of 11.170%, compared to 11.105% for intervening on the data and 11.158% for the base risk adjustment formula.[21]

Overall, in this application, intervening on the data is the most effective method for addressing underpayment for individuals with a mental illness within a balanced budget. Adding in risk adjustors is the least effective method. By design, intervening on the data and constrained regression are comparable in their results for the aggregate mental illness group, but intervening on the data moves closer to eliminating under and overpayment in all of the non-mental illness subgroups. Improvements in fit at the group level essentially come at no expense of fit at the individual level as measured by R^2.

These approaches to addressing underpayment also have differential impacts on subgroups within the mental illness group. Figure 1 compares changes in net compensation by method, across mental illness subgroups occurring in at least 10% of individuals with a mental illness. Intervening on the data substantially increases net compensation for the screening history of mental health and substance use disorders, anxiety disorders, and mood disorders groups. The dementia/delirium and alcohol/substance disorder groups have the highest average spending ($21,876/person and $19,766/person) – thus receiving the largest transfers – and become overpaid after we intervene on the data. The constrained regression performs similarly to intervening on the data in terms of the impact on mental illness subgroups. The regression that

[21] The baseline regression R^2 is slightly lower than that reported for the CMS-HCC Version 21 regression (12.46%). However, the correlation between predicted values from our regression and those using the CMS-HCC coefficients is 98%, indicating our empirical formula is a good representation of the one used by CMS for plan payment.

includes the RxHCCs for bipolar, depression, and major depression performs similarly to the baseline regression, although compared to constrained regression and intervening on the data it generates the largest increase in net compensation for individuals without a mental illness HCC. The formula with the added RxHCCs also substantially increases net compensation for the mood disorders subgroup, which includes depression and bipolar ICD-9 codes.

Figures 2 and 3 provide a side-by-side comparison of intervening on the data to specifically target individuals with mental illness (Figure 2) versus all individuals with at least one chronic condition (Figure 3). Figure 2 shows how increasing spending for individuals with a mental health illness and decreasing it for all others affects predicted spending – and thus net compensation – at the group level. Increasing spending by slightly less than 15% would eliminate underpayment for individuals with multiple chronic conditions and a mental illness, while the transfer increases the overpayment for the mental illness only group. Figure 3 shows intervening on the data to increase spending for individuals with multiple chronic conditions has a similar effect on the multiple chronic conditions groups with and without mental illness. Given the differing levels of initial net compensation, the multiple chronic conditions with mental illness group requires a spending transfer of approximately 12% while the multiple chronic conditions with no mental illness group requires only about a 2% increase to eliminate underpayment.

Reducing Disparities for Individuals in Low-Income Neighborhoods

The low-income group faces a disparity of $191 compared to the high-income (reference) group (Table 4). Examined by subgroups (zip code income quintiles), we see the disparity primarily occurs in the second and third income quintiles, and largely in relation to use in the fifth income quintile. The goal is to increase the average payment for the low-income group in order to decrease the disparity. Intervening on the data with a 10% spending transfer to the low-income group increases payments by approximately $25 per person (Table 5). The overall fit of the regression is minimally reduced from 11.158% by a 10% transfer: the refitted regression yields an R^2 of 11.114%. By contrast, adding in a risk adjustor for the low-income group has a large effect and makes average payments equal to average costs, as expected. The combination of adding a risk adjustor and transferring 10% leads to the most dramatic change in payments. The addition of the low-income group indicator means there is a risk adjustor in the formula that is perfectly imbalanced between the group of interest and the reference group. This risk adjustor provides a clear channel for intervening on the data to work through without greatly disturbing other coefficients in the risk adjustment formula.

6. Discussion

This paper contributes to the literature on health plan payment on both theoretical and practical levels. Our central theoretical insight is that there is a two-way relationship between the data describing the performance of the health care system and the level of payments that emerge from a risk adjustment methodology. Payments determine behavior, and behavior generates the data that feeds into payment algorithms. Our paper calls attention to the second link, and introduces a new mechanism for affecting payment and, thus, health system performance. The researcher/regulator need not accept the behavior from the flawed health care system as the input data for use in calibrating payment, and in general, they should not. The mechanism of intervening on the data is easy to implement in practice, and intuitive, and therefore may have readier practical application. Health care systems are complex and suffer from allocation problems related to efficiency and fairness. We show how intervening on the data is a simple and transparent way to address both types of problems.

We present two empirical applications, which assume the current health care system underspends on individuals with chronic illnesses and underprovides services for individuals living in low-income zip codes. In the context of net compensation by disease group, intervening on the data and constrained regression both outperform the approach of including additional risk adjustors. Both intervening on the data and constrained regression are effective methods, have little effect on R^2, and improve group fit. Adding in a risk adjustor for low-income groups exacerbates the health care disparity between low-income and high-income neighborhoods. Combining intervening on the data with adding a risk adjustor is more effective at altering incentives than intervening on the data in isolation.

Our goal in this paper is to introduce a new idea for improving the performance of health plan payment methods. While the empirical applications show how our model might be implemented, this method requires further practical development in terms of how it would be implemented with an existing health plan payment system. For example, policy objectives, groups to be targeted, and outcome measures would all need to be specified in each year of implementation. We use the Medicare program as the setting for our empirical demonstration, but in practice it may be more straightforward to implement intervening on the data in a health plan payment system where the data used for risk adjustment calibration is the same as the spending data in the current system the regulator seeks to affect. Both the Netherlands and Germany, for example, have such systems, and would be appropriate settings for further empirical work or policy applications.

Regulators may choose from a variety of policy alternatives to achieve an efficiency or fairness goal via the risk adjustment system. Expanding the formula with additional risk adjustors is, as discussed earlier in this paper, a common approach. However, it not only is harmful in the case of the disparities example, but it also has potential costs in an efficiency setting, such as upcoding or other types of gaming, which are avoided by intervening on the data. There is a very large set of ad hoc, post-estimation adjustments regulators can implement – i.e., increasing the prices paid for certain groups after estimating the risk adjustment formula – but these modifications do not attempt to preserve OLS fit properties and do not take into account other subgroups, whose coefficients/payments remain untouched. Our approach of intervening on the data is expands the regulator’s toolkit in order to facilitate achieving objectives that may be more costly or difficult when implementing other approaches like ad hoc adjustments or including more risk adjustors.

While the main objective of this paper has been to provide a new approach to payment system design, there is another important implication of the basic conceptual issue we explore here. It is not only problematic to use observed data to design payment systems, it is also problematic to use observed data to evaluate payment systems. R^2, predictive ratios, and under/overpayment measures of payment system performance evaluate how well plan revenues match observed costs. However, more appropriate measures of payment system performance would evaluate how well plan revenues match efficient levels of spending. Estimation and evaluation methods are clearly linked, and if a different estimation approach is warranted, different measures of performance are also likely needed.

A strand of the European literature on risk adjustment proposes to address this problem of using observed costs to evaluate risk adjustment by instead using “acceptable costs” or “normative expenditures” for payment algorithm evaluation (Stam, van Vliet and van de Ven 2010; van de Ven and Ellis 2000). Acceptable costs are those which capture medically necessary care delivered in a cost-effective manner. Our approach is conceptually similar – to set targets for acceptable levels of spending levels that reflect a social policy choice. We also explore a different implementation approach: rather than modifying the coefficients assigned to risk adjustment variables, we modify spending based on our beliefs about what acceptable costs should be for the groups we are concerned about.

“Getting the prices right” will be a central problem of health plan payment in health insurance markets around the world for years to come. There are many improvements that can be made to current methods. Our main argument here is that altering the data itself should be part of any policymaker’s toolkit.

References

Agency for Healthcare Research and Quality. "Clinical Classification Software (CCS) 2015." 2016.

Carey, Colleen. "Technological change and risk adjustment: benefit design in Medicare Part D." *American Economic Journal: Economic Policy* 9, no. 1 (2017): 38-73.

Centers for Medicare and Medicaid Services. "March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting: Discussion Paper." 2016.

Cook, Benjamin Le, Thomas G. McGuire, and Alan M. Zaslavsky. "Measuring Racial/Ethnic Disparities in Health Care: Methods and Practical Issues." *Health Services Research* 47, no. 3.2 (2012): 1232-1254.

Department of Health and Human Services. "Medicare Part D 2013 Model software/ICD-9 Mappings." 2017.

Frank, R.G., J. Glazer, and T.G. McGuire. "Measuring adverse selection in managed health care." *Journal of Health Economics* 19, no. 6 (2000): 829-854.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Regularization paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33, no. 1 (2010): 1-22.

Geruso, Michael, Timothy J. Layton, and Daniel Prinz. "Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges." *NBER Working Paper No. 22832*, 2016.

Glazer, Jacob, and Thomas G. McGuire. "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care." *American Economic Review* 90, no. 4 (2000): 1055-1071.

Glazer, Jacob, and Thomas G. McGuire. "Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment." *Journal of Public Economics* 84, no. 2 (2002): 153-173.

Han, Tony, and Kurt Lavetti. "Does Part D abet advantgeous selection in Medicare Advantage?" *Journal of Health Economics* 56 (2017): 368-382.

Hu, Lily, and Yiling Chen. "Fairness at equilibrium in the labor market." *Proceedings of Fairness, Accountability, and Transparency in Machine Learning.* Halifax, Nova Scotia, Canada, 2017. 5 pages.

Institute of Medicine. *Unequal Treament: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC.: National Academies Press, 2002.

Lavetti, Kurt J., and Kosali Simon. "Strategic formulary design in Medicare Part D Plans." *American Economic Journal: Economic Policy*, Forthcoming.

Layton, Timothy J., Randall P. Ellis, Thomas G. McGuire, and Richard van Kleef. "Measuring efficiency of health plan payment systems in managed competition health insurance markets." *Journal of Health Economics* 56 (2017): 237-255.

Layton, Timothy J., Thomas G. McGuire, and Richard C. van Kleef. "Deriving risk adjustment payment weights to maximize efficiency of health insurance markets." *NBER Working Paper No. 22642*, 2016.

Liaw, Andy, and Mattew Wiener. "Classification and Regression by randomForest." *R News* 2, no. 3 (2002): 18-22.

Lum, Kristian, and William Isaac. "To predict and serve?" *Significance* 13, no. 5 (2016): 14-19.

McGuire, Thomas G., and Joseph P. Newhouse. "Medicare Advantage: Regulated Competition in the Shadow of a Public Option." Chap. 19 in *Risk Adjustment, Risk Sharing, and Premium Regulation in Health Plan Payment: Theory and Practice*, edited by Thomas G. McGuire and Richard van Kleef. Elsevier, Forthcoming.

McGuire, Thomas G., Joseph P. Newhouse, Sharon-Lise Normand, Julie Shi, and Samuel Zuvekas. "Assessing incentives for service-level selection in private health insurnace exchanges." *Journal of Health Economics* 35 (2014): 47-63.

Montz, Ellen, Tim Layton, Alisa B. Busch, Randall P. Ellis, Sherri Rose, and Thomas G. McGuire. "Risk-Adjustment Simulations: Plans May Have Incentives to Distort Mental Health and Substance Use Coverage." *Health Affairs* 35, no. 6 (2016): 1022-1028.

Newhouse, Joseph P. *Pricing the Priceless.* The MIT Press, 2002.

Newhouse, Joseph P., Mary Price, John Hsu, J. Michael McWilliams, and Thomas G. McGuire. "How much favorable selection is left in Medicare Advantage?" *American Journal of Health Economics* 1, no. 1 (2015): 1-26.

Pearl, Judea. *Causality: Models, Reasoning and Inference.* New York: Cambridge University Press, 2009.

Petersen, Maya L., et al. "Super Learner Analysis of Electronic Adherence Data Improves Viral Prediction and May Provide Strategies for Selective HIV RNA Monitoring." *Journal of Acquired Immune Deficiency Syndromes* 69, no. 1 (2015): 109-118.

Pirracchio, Romain, Maya L. Petersen, Marco Carone, Matthieu Resche Rigon, Sylvie Chevret, and Mark J. van der Laan. "Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study." *The Lancet: Respiratory Medicine* 3, no. 1 (January 2015): 42-52.

Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. *Evaluation of the CMS-HCC Risk Adjustment Model: Final Report.* Centers for Medicare and Medicaid Services, 2011.

Rose, Sherri. "A Machine Learning Framework for Plan Payment Risk Adjustment." *Health Services Research* 51, no. 6 (2016): 2358-2374.

Rose, Sherri, Savannah L. Bergquist, and Timothy J. Layton. "Computational health economics for identifying unprofitable health care enrollees." *Biostatistics* 18, no. 4 (2017): 682-694.

Shepard, Mark. "Hospital network competition and adverse selection: Evidence form the Massachusetts health insurance exchange." *NBER Working Paper* No. 22600 (2016).

Stam, Pieter J.A., R.C.J.A van Vliet, and Wynand P.M.M. van de Ven. "A limited-sample benchmark approach to assess and improve the performance of risk equalization models." *Journal of Health Economics* 29, no. 3 (2010): 426-437.

Van de Ven, Wynand P.M.M., and Randall P. Ellis. *Chapter 14 - Risk Adjustment in Competitive Helath Insurance Plan Markets.* Vol. 1, in *Handbook of Health Economics*, by Anthony J. Culyer and Joseph P. Newhouse, 755-845. Elsevier, 2000.

Van der Laan, Mark J., and Sherri Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data.* New York: Springer, 2011.

Van der Laan, Mark, Eric C. Polley, and Alan E. Hubbard. "Super Learner." *Statistical Applications in Genetics and Molecular Biology* 6, no. 1 (2007).

Van Kleef, Richard C., Rene C.J.A van Vliet, and Wynand P.M.M. van de Ven. "Risk equalization in the Netherlands: an empirical evaluation." *Expert Reviews Pharacoeconomics Outcomes Research* 13, no. 6 (2013): 829-839.

Van Kleef, Richard, Thomas McGuire, Rene van Vliet, and Wynand van de Ven. "Improving risk equalization with constrained regression." *The European Journal of Health Economics* 18, no. 9 (2017): 1137-1156.

Van Veen, S.C.H.M., R.C. van Kleef, W.P.M.M. van de Ven, and R.C.J.A. van Vliet. "Is there one measure-of-fit that fits all? A taxonomy and review of measures-of-fit for risk-equalization models." *Medical Care Research and Review* 72, no. 2 (2015): 220-243.

Tables and Figures

Table 1: Sample Characteristics in 2010 (N=1.5 million)

Variable	Mean (%)
Sex	
Female	55
Age	
<65	17
65-69	21
70-74	17
75-79	16
80+	25
Race	
White	85
Black	9
Hispanic	6
Urban/Rural	
Rural	24
Region	
Northeast	18
Midwest	24
South	40
West	17
Chronic Conditions	
Mental illness	22
Cancer	23
Heart disease	41
Diabetes	27
At least one chronic condition	68
Insurance	
Medicaid	18

Table 2: Net Compensation and Average Spending (in U.S. Dollars, 2011) by Mutually Exclusive Chronic Condition Groups as Defined by CCS Categories

Group	**Net Compensation**	**SE**	**Mean Spending**	**SE**	**N**
Mental illness	-652	66	15,671	70	335,592
Mental illness only	315	64	7,580	66	98,130
Multiple chronic conditions, mental illness	-1,052	89	19,014	94	237,462
Multiple chronic conditions, no mental illness	-92	72	14,623	75	266,565
Diabetes only	768	55	6,687	56	117,257
Heart disease only	-66	64	10,018	66	197,373
Cancer only	247	62	7,740	64	106,672
No chronic condition	294	24	4,416	24	476,541

Table 3: Comparison of Methods Targeting Mental Illness Group – Net Compensation by Mutually Exclusive Groups (in U.S. Dollars, 2011)

Group	**Baseline OLS**	**Intervening on the Data**		**Adding Risk Adjustors**	**Constrained Regression**		**N**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	
GPSF	92.032	98.009	95.259	93.262	96.650	96.777	1,500,000
R^2	11.158	11.105	11.127	11.170	11.127	11.125	1,500,000
Mental illness only	315	644	480	518	988	1,006	98,130
Multiple chronic conditions, mental illness	-1052	-291	-672	-825	-433	-416	237,462
Multiple chronic conditions, no mental illness	-92	-31	-62	-171	-214	-218	266,565
Diabetes only	768	519	643	702	563	557	117,257
Heart disease only	-66	-176	-121	-133	-227	-232	197,373
Cancer only	247	31	139	195	33	27	106,672
No chronic conditions	294	-32	131	239	80	75	476,451

Notes: Column (1) contains the simplified CMS-HCC Version 21 risk adjustment model. (2) contains intervening on the data where we increase spending by 10% for all persons with a mental illness. (3) contains intervening on the data where we increase spending by 5% for all persons with a chronic condition. (4) is the baseline OLS model with three additional RxHCC risk adjustor indicators. (5) is a constrained regression with a constraint of $\bar{Y}_{MH} - 17.41 = \sum_j \beta_j \, \bar{z}_{MH,j}$ (targeting the average spending for the mental illness group achieved by the intervening on the data implementation in (2). (6) is a constrained regression with a constraint of $\bar{Y}_{MH} = \sum_j \beta_j \, \bar{z}_{MH,j}$ to achieve zero net compensation for the mental illness group.

Table 4: Disparity, Initial Mean Payments, and Mean Spending by Neighborhood Income Groups (in U.S. Dollars, 2011)

Group	Disparity	Mean Payments	Mean Spending	N
Zip Code Income				1,500,000
Binary				
Low	-191	10,050	9,911	900,000
High	0	9,673	9,880	600,000
Quintiles				
1st	-58	10,227	10,160	300,000
2nd	-251	10,051	9,852	300,000
3rd	-265	9,872	9,723	300,000
4th	0	9,786	9,760	300,000
5th	0	9,559	10,001	300,000

Table 5: Comparison for Reducing Disparities by Neighborhood Income – Group Payments and Spending (in U.S. Dollars, 2011)

Group	Base OLS	Intervening on the Data	Adding a Risk Adjustor	Combination	Mean Spending
	(1)	(2)	(3)	(4)	(5)
R^2	11.158	11.114	11.161	11.094	
Zip Code Income					
Binary					
Low	10,050	10, 074	9,911	10,308	9,911
High	9,673	9,636	9,880	9,286	9,880
Quintiles					
1st	10,227	10,284	10,095	10,506	10,160
2nd	10,051	10,071	9,911	10,307	9,852
3rd	9,872	9,868	9,728	10,110	9,723
4th	9,786	9,765	9,996	9,410	9,760
5th	9,559	9,508	9,765	9,161	10,001

Notes: Column (1) contains the simplified CMS-HCC v21 risk adjustment regression. (2) contains intervening on the data where we increase spending by 10% for individuals in the low-income group. (3) is the baseline OLS regression with an indicator the low-income group. (4) contains the baseline OLS regression with an indicator for the low-income group, combined with intervening on the data where we increase spending by 10% for the low-income group. (5) shows mean group-level spending.

Figure 1: Comparing Changes to Net Compensation for Mental Illness Subgroups by Methodological Approach

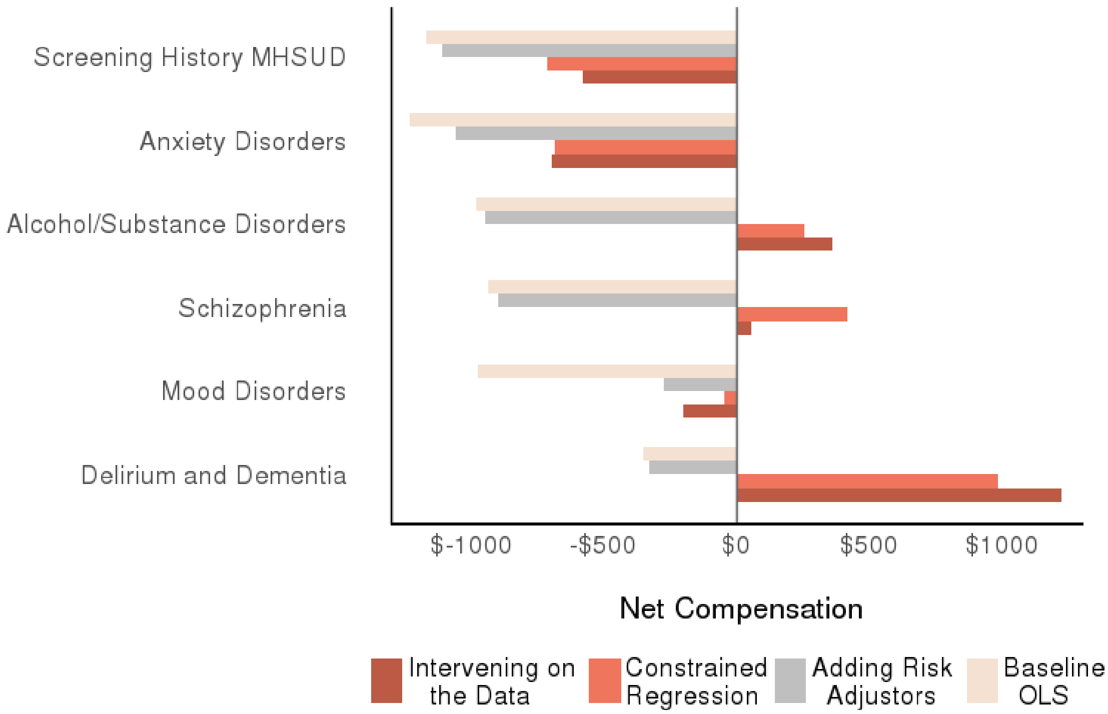


Figure 2: Relationship between Intervening on the Data Spending Transfer for Mental Illness Group and Net Compensation by Chronic Condition

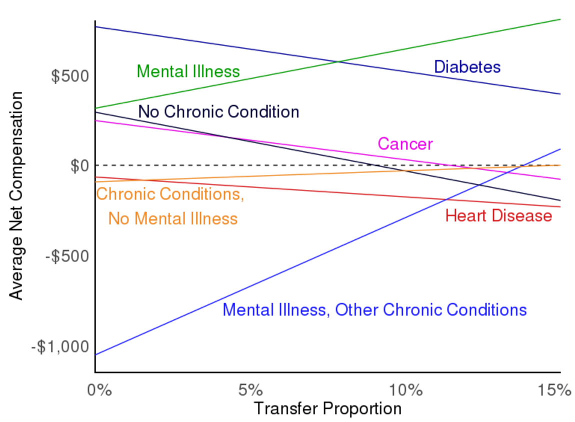
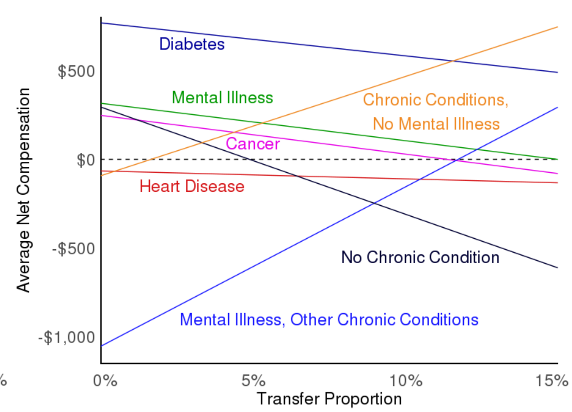


Figure 3: Relationship between Intervening on the Data Spending Transfer for Chronic Condition Group and Net Compensation by Chronic Condition



Appendix

Baseline Risk Adjustment Formula

As described in the main text, the baseline risk adjustment regression we implement is a similar specification to the Version 21 CMS-HCC formula for the aged, community dwelling subgroup. We simplify our formula and do not include interaction terms between diseases, which are sparsely populated, and we also do not modify coefficients post-estimation as CMS does (Pope, et al. 2011). To calculate risk scores, CMS converts the dollar risk adjustment coefficients into relative risk factors by dividing by the average per person predicted spending for a designated year. In our analyses we use the coefficients as dollar weights. With an analysis for one year, these methods are equivalent.

Super Learner

Machine learning is well-suited to estimating disparities because our definition is formulated as a prediction exercise and the functional form of any differences is not known in advance. We therefore adjust for health status in a general and flexible way, using ensemble machine learning methods. We use the super learner framework, which combines many algorithms into a weighted average prediction function (van der Laan, Polley and Hubbard 2007, van der Laan and Rose 2011). While it has not, to our knowledge, been used in disparities work, it has been deployed for prediction or classification in other applications, such as health outcomes (e.g., Petersen, et al. 2015 and Pirracchio, et al. 2015) and health spending (e.g., Rose 2016 and Rose, Bergquist and Layton 2017).

The super learner takes as input: (1) the data and (2) a collection of algorithms. The first step involves performing V-fold cross-validation for each algorithm in the collection. An optimal weight vector is then constructed by regressing the outcome Y_i on the cross-validated predicted spending values generated for each algorithm. The estimated coefficient for each column in the matrix of cross-validated predicted spending values is the weight for the algorithm that produced those values. One can show that this regression selects the optimal weight vector, and that by restricting the family of possible weighted combinations to convex combinations (i.e., nonnegative weights that sum to one) performance can be improved (van der Laan, Polley, and Hubbard 2007). The collection of algorithms the super learner ensembles can include not only different classes of algorithms (e.g., decision trees and penalized regressions), but also multiple similar algorithms with alternative tuning parameters (e.g., random forests with 250 observations in terminal nodes vs. 500 observations in terminal nodes). Algorithms, such as penalized regressions, that may select tuning parameters (e.g., λ), via cross-validation will still do so within the super learner procedure, leading to nested layers of cross-validation. The final super learner algorithm fits each algorithm on the full data and combines those functions with the estimated optimal weights to produce a final predicted value. Our super learner considered six algorithms: a main terms logistic regression, three penalized regressions, and a random forest with 500 trees and a minimum node size of 250 observations. We implement our analyses using the SuperLearner package in R, which calls the randomForest and glmnet packages (Liaw and Wiener 2002; Friedman, Hastie and Tibshirani 2010).

Additional risk adjustors

We add three risk adjustors from the Medicare Part D risk adjustment formula (CMS-RxHCC), which has been shown to recognize a greater number of individuals with mental health conditions in the ACA Marketplace setting (Montz, et al. 2016). The goal is to choose additional risk adjustors to capture individuals in the mental illness group (as defined by CCS groups) not already well represented by the included HCCs. Appendix Table 1 shows the overlap in underlying ICD-9 codes for each mental illness CCS group and the corresponding HCCs and RxHCCs. We select indicators for bipolar disorders, major depression, and depression to include in our regression. To pick these three risk adjustors, we first examined RxHCCs with the fewest ICD-9 codes in common with the mental illness HCCs, and then we chose the RxHCCs that covered the largest number of individuals in the CCS mental illness group.

Appendix Table 1: CCS ICD9 Codes cross walk with HCCs and RxHCCs

ICD9 Codes	CCS Group	HCC	RxHCC
3090 3091 30922 30923 30924 30928 30929 3093 3094 30982 30983 30989 3099	Adjustment disorders (650)	N/A	Depression (62)
29384 30000 30001 30002 30009 30010 30020 30021 30022 30023 30029 3003 3005 30089 3009 3080 3081 3082 3083 3084 3089 30981 3130 3131 31321 31322 3133 31382 31383	Anxiety Disorders (651)	N/A	Anxiety Disorders (63) Specified Anxiety, Personality, and Behavior Disorders (61)
31200 31201 31202 31203 31210 31211 31212 31213 31220 31221 31222 31223 3124 3128 31281 31282 31289 3129 31381 31400 31401 3141 3142 3148 3149	ADHD, conduct, and disruptive behavior disorders (652)	N/A	Specified Anxiety, Personality, and Behavior Disorders (61)
2900 29010 29011 29012 29013 29020 29021 2903 29040 29041 29042 29043 2908 2909 2930 2931 2940 2941 29410 29411 29420 29421 2948 2949 3100 3102 3108 31081 31089 3109 3310 3311 33111 33119 3312 33182 797	Delirium, dementia, and amnestic and other cognitive disorders (653)	Dementia without Complication (52) Dementia with Complications (51)	Dementia, Except Alzheimer’s Disease (55) Alzheimer’s Disease (54)
3070 3079 31500 31501 31502 31509 3151 3152 31531 31532 31534 31535 31539 3154 3155 3158 3159 317 3180 3181 3182 319 V400 V401	Developmental disorders (654)	N/A	Mild or Unspecified Mental Retardation/ Developmental Disability (68) Profound or Severe Mental Retardation/ Developmental Disability (66)

			Moderate Mental Retardation/ Developmental Disability (67)
29900 29901 29910 29911 29980 29981 29990 29991 30720 30721 30722 30723 3073 3076 3077 30921 31323 31389 3139	Disorders usually diagnosed in infancy, childhood, or adolescence (655)	N/A	Specified Anxiety, Personality, and Behavior Disorders (61) Autism (65)
31230 31231 31232 31233 31234 31235 31239	Impulse control disorders, NEC (656)	N/A	Specified Anxiety, Personality, and Behavior Disorders (61)
29383 29600 29601 29602 29603 29604 29605 29606 29610 29611 29612 29613 29614 29615 29616 29620 29621 29622 29623 29624 29625 29626 29630 29631 29632 29633 29634 29635 29636 29640 29641 29642 29643 29644 29645 29646 29650 29651 29652 29653 29654 29655 29656 29660 29661 29662 29663 29664 29665 29666 2967 29680 29681 29682 29689 29690 29699 3004 311	Mood disorders (657)	Major Depressive, Bipolar, and Paranoid Disorders (58)	Bipolar Disorders (59) Major Depression (60) Depression (62)
3010 30110 30111 30112 30113 30120 30121 30122 3013 3014 30150 30151 30159 3016 3017 30181 30182 30183 30184 30189 3019	Personality disorders (658)	N/A	Depression (62) Specified Anxiety, Personality, and Behavior Disorders (61)
29381 29382 29500 29501 29502 29503 29504 29505 29510 29511 29512 29513 29514 29515 29520 29521 29522 29523 29524 29525 29530 29531 29532 29533 29534 29535 29540 29541	Schizophrenia and other psychotic disorders (659)	Schizophrenia (57) Major Depressive, Bipolar, and Paranoid Disorders (58)	Schizophrenia (58)

29542 29543 29544 29545 29550 29551 29552 29553 29554 29555 29560 29561 29562 29563 29564 29565 29570 29571 29572 29573 29574 29575 29580 29581 29582 29583 29584 29585 29590 29591 29592 29593 29594 29595 2970 2971 2972 2973 2978 2979 2980 2981 2982 2983 2984 2988 2989			
2910 2911 2912 2913 2914 2915 2918 29181 29182 29189 2919 30300 30301 30302 30303 30390 30391 30392 30393 30500 30501 30502 30503 3575 4255 5353 53530 53531 5710 5711 5712 5713 76071 9800	Alcohol-related disorders (660)	Drug/Alcohol Psychosis (54) Drug/Alcohol Dependence (55) Polyneuropathy (75) Congestive Heart Failure (85) Cirrhosis of Liver (28)	Polyneuropathy (74) Congestive Heart Failure (87)
2920 29211 29212 2922 29281 29282 29283 29284 29285 29289 2929 30400 30401 30402 30403 30410 30411 30412 30413 30420 30421 30422 30423 30430 30431 30432 30433 30440 30441 30442 30443 30450 30451 30452 30453 30460 30461 30462 30463 30470 30471 30472 30473 30480 30481 30482 30483 30490 30491 30492 30493 30520 30521 30522 30523 30530 30531 30532 30533 30540 30541 30542 30543 30550 30551 30552 30553 30560 30561 30562 30563 30570 30571 30572 30573 30580 30581 30582 30583 30590 30591 30592 30593 64830 64831 64832 64833 64834 65550 65551 65553 76072 76073 76075 7795 96500 96501 96502 96509 V6542	Substance-related disorders (661)	Drug/Alcohol Psychosis (54) Drug/Alcohol Dependence (55)	N/A
E9500 E9501 E9502 E9503 E9504 E9505 E9506 E9507	Suicide and	N/A	N/A

E9508 E9509 E9510 E9511 E9518 E9520 E9521 E9528 E9529 E9530 E9531 E9538 E9539 E954 E9550 E9551 E9552 E9553 E9554 E9555 E9556 E9557 E9559 E956 E9570 E9571 E9572 E9579 E9580 E9581 E9582 E9583 E9584 E9585 E9586 E9587 E9588 E9589 E959 V6284	intentional self-inflicted injury (662)		
3051 30510 30511 30512 30513 33392 7903 V110 V111 V112 V113 V114 V118 V119 V154 V1541 V1542 V1549 V1582 V6285 V663 V701 V702 V7101 V7102 V7109 V790 V791 V792 V793 V798 V799	Screening and history of mental health and substance abuse codes (663)	N/A	N/A
29389 2939 30011 30012 30013 30014 30015 30016 30019 3006 3007 30081 30082 3021 3022 3023 3024 30250 30251 30252 30253 3026 30270 30271 30272 30273 30274 30275 30276 30279 30281 30282 30283 30284 30285 30289 3029 3060 3061 3062 3063 3064 30650 30651 30652 30653 30659 3066 3067 3068 3069 3071 30740 30741 30742 30743 30744 30745 30746 30747 30748 30749 30750 30751 30752 30753 30754 30759 30780 30781 30789 3101 316 64840 64841 64842 64843 64844 V402 V403 V4031 V4039 V409 V673	Miscellaneous mental health disorders (670)	N/A	Specified Anxiety, Personality, and Behavior Disorders (61) Anxiety Disorders (63)

Disparities Application Extension

We expand on the primary disparities example by demonstrating an even more targeted intervention: once again 10% is transferred to the low-income group, but funds are taken from only the upper half of the high-income group (the fifth quintile), leaving individuals in the fourth income quintile unaffected. To accomplish this we modify the *Y* vector as before for the low-income group, imposing an overall budget constraint, and imposing a second constraint that maintains the spending levels of the fourth quintile, so that only individuals in the fifth quintile are used to counterbalance the spending for the low-income group.

Appendix Table 2 shows transferring 10% of spending to the low-income group but reducing spending for only the fifth income quintile generates minimal change in payments for the low-income group. Adding a risk adjustor for the low-income group does not allow us to specify what happens within the high-income group, and we see the fourth income quintile receives substantially higher payments. Combining adding a risk adjustor with intervening on the data is once again the most powerful approach, although the overall changes are much more modest (from $10,050 to $10,058 for the low-income group) than when we allowed spending for the entire high-income group to be reduced.

Appendix Table 2: Comparison for Reducing Disparities by Targeting Neighborhood Income Subgroup – Group Payments and Spending (in U.S. Dollars, 2011)

Group	**Base OLS**	**Intervening on the Data**	**Adding a Risk Adjustor**	**Combination**	**Mean Spending**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
R^2	11.158	11.117	11.161	11.116	
Zip Code Income					
Binary					
Low	10,050	10,053	9,911	10,058	9,911
High	9,673	9,668	9,880	9,660	9,880
Quintiles					
1st	10,227	10,230	10,095	10,261	10,160
2nd	10,051	10,057	9,911	10,057	9,852
3rd	9,872	9,873	9,728	9,858	9,723
4th	9,786	9,786	9,996	9,786	9,760
5th	9,559	9,550	9,765	9,534	10,001

Notes: Column (1) contains the simplified CMS-HCC v21 risk adjustment regression. (2) contains intervening on the data where we increase spending by 10% for individuals in the low-income group and reduce spending for the 5th quintile. (3) is the baseline OLS regression with an indicator the low-income group. (4) contains the baseline OLS regression with an indicator for the low-income group, combined with intervening on the data where we increase spending by 10% for the low-income group and reduce spending for the 5th quintile. (5) shows mean group-level spending.