Measuring Changes in Disparity Gaps: An Application to Health Insurance

By Paul Goldsmith-Pinkham, Karen Jiang, Zirui Song, and Jacob Wallace[*]

It is now widely recognized that achieving health equity for members of racial and ethnic minority groups requires isolating and intervening on the manipulable factors that underlie health inequities (Bailey et al., 2017). The key empirical challenge—identifying the sources of differences across groups (typically means)—echoes work in other contexts outside of health—e.g., the gender wage gap (Blau and Kahn, 2017); or the racial wealth gap (Derenoncourt et al., 2021). One approach in this literature is the Kitagawa-Oaxaca-Blinder approach (Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973), which measures how much mean differences across groups in relevant covariates can explain the gaps.

While the Kitagawa-Oaxaca-Blinder (KOB) approach has been used to decompose mean differences between groups in observational settings (Fortin, Lemieux and Firpo, 2011) or as a tool for estimating treatment effects (Kline, 2011; Słoczyński, 2015), in this paper we show that KOB-style decompositions can also be used to decompose sources of treatment effect heterogeneity in policies that reduce disparities, with a focus on identifying sources of treatment effect heterogeneity between groups. Our approach is closely related to work on decomposing sources of treatment effect variation (Heckman, Smith and Clements, 1997; Djebbari and Smith, 2008; Ding, Feller and Miratrix, 2019; Currie, Voorheis and Walker, 2020), but also to a more recent literature on how to estimate heterogeneity in causal effects (e.g., Athey and Imbens, 2016; Wager and Athey, 2018; Denteh et al., 2021).

[*] Goldsmith-Pinkham: Yale School of Management, 165 Whitney Avenue, New Haven, CT 06511, e-mail: paul.goldsmith-pinkham@yale.edu. Jiang: Yale University, PO Box 208421, New Haven, CT 06520. e-mail: karen.jiang@yale.edu. Song: Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115. e-mail: Song@hcp.med.harvard.edu. Wallace: Yale School of Public Health, 60 College Street, New Haven, CT 06520. e-mail: jacob.wallace@yale.edu.

We first show that the reduction in disparities between groups caused by a treatment can be written as the difference in conditional average treatment effects (CATEs) for each group. Then, we highlight that these CATEs can be decomposed into CATEs driven by other observables (e.g., the "endowment" difference) and unexplained differences between groups. We argue that reporting the share driven by each component can be an important summary statistic for researchers interested in understanding group differences, since it forces researchers to focus on characteristics that are causally manipulable, rather than features like gender or race. Finally, we apply this approach to study the impact of Medicare on racial and ethnic disparities in healthcare access and outcomes.

I. Heterogeneous Treatment Effects and Disparity Gaps

Consider a population of individuals with randomly assigned treatment $D_i \in \{0, 1\}$, a discrete multivalued set of controls X_i and a set of potential outcomes $Y_i(1)$ and $Y_i(0)$, with $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. We define a binary group indicator W_i such as race or gender. We assume strong ignorability:

$$(1) \qquad (Y_i(0), Y_i(1), X_i, W_i) \perp D_i$$

Researchers are often estimate the effect of the treatment D_i for each group W_i and compare the treatment effect heterogeneity for each group. We first show that the difference in treatment effects across group W can be written as the difference in the *outcome gap* across W_i before and after the treatment. In other words, the difference in treatment effects across groups is exactly the change in disparity gaps.

Let $\mu_d(w) = E(Y_i | D_i = d, W_i = w)$, and $\mu_d(w, x) = E(Y_i | D_i = d, W_i = w, X_i = x)$. Then, $\tau = \mu_1 - \mu_0$, $\tau(w) = \mu_1(w) - \mu_0(w)$ and $\tau(w, x) = \mu_1(w, x) - \mu_0(w, x)$. Here, τ identi-

fies the ATE and $\tau(w)$ and $\tau(w,x)$ each identify different conditional average treatment effects (CATE), where we assume overlap is satisfied. Finally, let $\gamma_d = \mu_d(1) - \mu_d(0)$ be the disparity gap between the two groups in either the control or treatment case.

Then, we note that the change in the disparity gap is exactly the difference in treatment effects:

$$(2) \qquad \delta = \gamma_1 - \gamma_0 = \tau(1) - \tau(0).$$

Rewriting group differences in this way is powerful. Since most researchers consider W_i which are not causally manipulable (e.g. race or gender) under the "no causation without manipulation" maxim of Holland (1986), γ_d is not a well-defined causal object. But, τ_w is a meaningful causal object with clear interpretation. When comparing τ_1 and τ_0, the change in the gap between the groups can be re-interpreted as the difference in treatment effects, such that decomposing how the heterogeneity affects disparity gaps can be informative about the source of the disparities themselves.

Next, we can rewrite this change in the overall gap as a combination of two pieces, analogous to a Kitagawa-Oaxaca-Blinder (KOB) decomposition: a weighted combination of different heterogeneous effects within characteristics, plus an adjustment given a difference in composition of characteristics x. Let $\pi_w(x) = Pr(X_i = x | W_i = w)$ and consider the following decomposition:

$$(3) \qquad \delta = \sum_x \tau(1,x)\pi_1(x) - \tau(0,x)\pi_0(x)$$

$$(4) \qquad = \sum_x \delta(x)\pi_1(x)$$

$$(5) \qquad \quad + \sum_x \tau(0,x)(\pi_1(x) - \pi_0(x)).$$

Note that these are two types of disparity gaps: one is a difference in outcomes *given* a characteristic, and the other is a gap due to differences across groups in observable characteristics. Since $\delta(x)$ can be defined as either a difference in treatment effects, or a change in disparity gaps following a treatment, it is possible to interpret the difference in treatment effects as a change in disparities within characteristics (weighted by one group's distribution of characteristics) and the pre-existing difference in characteristics that affect treatment (weighted by one group's treatment effect). An important caveat is that alternative weightings are feasible based on whether the analysis uses $w = 1$ or $w = 0$ as the reference group (as with KOB decompositions). Researchers are encouraged to confirm that the implication of the results below do not shift meaningfully depending on the choice of reference.

A natural measure for this heterogeneity is $\kappa = \sum_x \delta(x)\pi_1(x)/\delta$. κ can be larger than 1 or less than 0. Two extreme cases are worth noting:

1) $\tilde{\delta} = 0$. Then, all effects are driven by cross-characteristics composition: $\delta = \sum_x \tau(x)(\pi_1(x) - \pi_0(x))$, and $\kappa = 0$.

2) $\pi_1(x) - \pi_0(x) = 0$ for all x. Then, $\delta = \sum_x \delta(x)Pr(X_i = x)$, and all effects are driven by within-characteristic heterogeneity differences. Then, $\kappa = 1$.

There are two important things to note about use of this decomposition. First, if a policymaker considers targeting the treatment, a large κ suggests that benefits accrue *within*-characteristics. As a result, policy makers will achieve reductions in disparity gaps by ensuring take-up of the treatment within a given place where $\delta(x)$ is large. Of course, focusing on disparities may ignore cases where both $\tau_1(x)$ and $\tau_0(x)$ are large, but the difference $\delta(x)$ is small. In contrast, if κ is small, disparities can be reduced by targeting the treatment to where the covariance between $\tau(x)$ and $\pi_1(x) - \pi_0(x)$ is positive.

Second, consider the setting when $\kappa = 0$. In this setting, is it possible to say that differences in x cause the initial disparity gap? In other words, if we causally manipulated X such that the distribution of X was identical across W, how would this change γ_0? Without additional assumptions, this is not knowable. If individuals sort by X, then our estimates of $E(Y|X_i = x)$ do not identify the counterfactual average for all individuals.

II. Data and Empirical Framework

We use this approach to study the impact of eligibility for Medicare—(nearly) universal public health insurance that most Americans qualify for beginning at age 65—on racial and ethnic disparities in healthcare access and outcomes.

The data used in the study come from the Behavioral Risk Factor Surveillance System (BRFSS) for the period 2010-2018, a health-related telephone survey that collects demographics and health-related outcomes from individuals in all 50 states (United States Department of Health and Human Services. Centers for Disease Control and Prevention, 2020). See the Online Appendix for additional details on the BRFSS. Importantly, the data include age in years and state identifiers, allowing us to estimate the local average treatment effect of eligibility for Medicare at age 65 in each region of the country using a regression discontinuity design (Goldsmith-Pinkham, Pinkovskiy and Wallace, 2020; Wallace et al., 2021).[1]

In our application, we limit the data to respondents aged 51-79 years and estimate regression discontinuity analyses (see the Online Appendix for additional details on the estimation) both at the national and region levels, and separately by race/ethnicity, using equations of the form:

$$
(6) \quad \begin{aligned}
y_{i,j,l}(\text{age}) = {} & \tau \times 1(\text{age} > 65) \\
& + f(\text{age}) \times 1(\text{age} \le 65) \\
& + g(\text{age}) \times 1(\text{age} > 65) \\
& + \varepsilon_{i,t,l}(\text{age}).
\end{aligned}
$$

where $y_{i,j,l}(\text{age})$ is an outcome for individuals i of type j (i.e., self-identified race/ethnicity) in location l of a given age. The functions $f(\text{age})$ and $g(\text{age})$ are the age profile of $y_{i,j,l}$ for those below and above age 65, respectively, and adjust for the effect of age on our outcomes. The two outcomes to illustrate our approach are the share of survey respondents that report being unable to see a physician in the past year due to cost and the share of respondents with any source of health insurance coverage. We report additional outcomes in the Online Appendix.

Using this framework, we can estimate τ, the treatment effect of Medicare (D_i), for each outcome. Likewise, for each outcome we can also estimate $\tau(w,x)$ and $\tau(w)$ by re-estimating the RD regression within those groups. This provides the necessary pieces to estimate δ and κ.

---

[1]However, in principle these decomposition methods could be used in the context of program evaluations using a range of research designs (e.g., randomized controlled trials, difference-in-differences, etc.).

III. Results

An advantage of the RD approach used in this application is that it allows us to easily visualize the treatment effects γ graphically, as well as the extent to which the treatment reduces racial and ethnic disparities at age 65. In Figure 1 of the Online Appendix, we report plots of the RD estimates separately by race/ethnicity and for states grouped into three groups: "All States", "Southern States" and "Other States." We see clear evidence of pre-Medicare disparities, such as in Panel A, where prior to being eligible for Medicare at age 65, there are visible racial and ethnic differences in the share of respondents that report being unable to see a physician in the past year due to cost, with the rates being highest for Hispanic respondents (27.3%) in Southern States and lowest for non-Hispanic, White respondents (7.7%) in Other States.

At the national level, we see a discontinuous reduction at age 65 in the Hispanic-White gap in the share of respondents unable to see a physician due to cost (Table 1). As we already noted above, the decline in the Hispanic-White gap (4.0pp) can be expressed as the difference in the CATEs for Hispanic respondents (7.0pp) and non-Hispanic White respondents (3.0pp). Relative to the pre-65 gap of 17.9 p.p., this represented a 25% reduction in the disparity.

We then perform a KOB-style decomposition using our region-level estimates to assess how much of the national-level reduction in the Hispanic-White gap is driven by a reduction in the gap within each region (i.e., "Southern States" vs. "Other States") at age 65 and how much is driven by treatment effect heterogeneity across regions and differences in where racial and ethnic groups reside (e.g., Black Americans are concentrated geographically in the South). In principle, this decomposition could be performed at a more granular level (e.g., states or counties) but for illustrative purposes we focus on two regions in this application.

For each of the outcomes we examined, the share of the national disparity reduction driven by within-region reductions in racial and ethnic disparities, κ, was large. For example, these estimates ranged from 0.79 for the Black-White gap in the share of the population without a usual source of care to approximately 1.0 for the Hispanic-White gap in all four outcomes.

Table 1—: Changes in Racial and Ethnic Gaps and the Effects of Medicare Eligibility at Age 65

Share unable to see physician in the past year due to cost	Share of Change in Gap Explained (κ)		Estimated Effect (τ)			Change in Gap (δ)	
	Black Americans	Hispanic Americans	White Americans	Black Americans	Hispanic Americans	Black Americans	Hispanic Americans
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Overall			-0.03	-0.05	-0.07	-0.02	-0.04
			(0.003)	(0.01)	(0.01)	(0.01)	(0.015)
Breakdown by Region	0.97	1.0					
Non-South			-0.03	-0.04	-0.03	-0.01	-0.01
			(0.004)	(0.01)	(0.01)	(0.01)	(0.02)
South			-0.03	-0.05	-0.12	-0.02	-0.09
			(0.005)	(0.02)	(0.03)	(0.02)	(0.04)
Share Insured							
Overall			0.063	0.092	0.15	0.029	0.083
			(0.003)	(0.01)	(0.01)	(0.01)	(0.015)
Breakdown by Region	0.85	0.99					
Non-South			0.056	0.084	0.1	0.028	0.047
			(0.003)	(0.01)	(0.02)	(0.02)	(0.02)
South			0.076	0.097	0.21	0.021	0.14
			(0.005)	(0.01)	(0.03)	(0.02)	(0.03)

Source: Authors analysis of the Behavioral Risk Factor Surveillance System (BRFSS), 2010-2018. This table reports estimated effects from the Medicare RD regression, for White, Black, and Hispanic Americans. Columns 1 and 2 explain the share of the change in the gap explained by heterogeneity within covariates (see text for details). Columns 3, 4 and 5 report estimates of the RD regression (equation 6) by racial and ethnic group. Columns 6 and 7 report the difference between columns 4 and 3, and 5 and 3, respectively. These can be interpreted as the discontinuity in the disparity at age 65.

The, perhaps surprising, implication of this is that the reductions in disparities at the national-level are not the result of Medicare having larger effects in a particular region (e.g., the South) where members of racial and ethnic minority groups are more likely to reside. Instead, it is driven by large reductions in the Black-White and Hispanic-White gaps *within regions* at age 65.

IV. Conclusion

We show that differences in treatment effects across groups can be interpreted as changes in disparity gaps. Moreover, we show the KOB approach can be extended to decompose differences in disparity gaps into differences in the observable characteristics correlated with treatment effect heterogeneity (e.g., region in our example) and treatment effect heterogeneity between groups (e.g., non-Hispanic White vs. Hispanic respondents) within a region.

This method may be particularly useful for researchers focused on understanding how treatment effect heterogeneity leads some policies to be disparity-reducing (e.g., Medicare eligibility) and interested in decomposing the sources of treatment effect variation between groups (e.g., by race, gender, and their intersections). If observable characteristics only explain a limited amount of the heterogeneity in the effect of a policy, that suggests there may exist important (currently unobserved) characteristics that differ by group. Researchers should attempt to identify these to better understand the disparate impacts of policies between groups. Researchers may want to split by more characteristics, potentially introducing additional noise. Further work should explore the use of machine techniques such as causal forests to implement this (Wager and Athey, 2018).

REFERENCES

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Bailey, Zinzi D, Nancy Krieger, Madina Agénor, Jasmine Graves, Natalia Linos, and Mary T Bassett. 2017. "Structural racism and health inequities in the USA: evidence and interventions." *The Lancet*, 389(10077): 1453–1463.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of economic literature*, 55(3): 789–865.

Blinder, Alan S. 1973. "Wage discrimination: reduced form and structural estimates." *Journal of Human resources*, 436–455.

Currie, Janet, John Voorheis, and Reed Walker. 2020. "What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality." National Bureau of Economic Research.

Denteh, Augustine, Tim Layton, Jacob Wallace, and Sherri Rose. 2021. "Using Machine Learning to Estimate the Heterogeneous Impacts of Medicaid Managed Care." *Manuscript, Tulane University*.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2021. "The racial wealth gap, 1860-2020." *Manuscript, Princeton University and University of Bonn*.

Ding, Peng, Avi Feller, and Luke Miratrix. 2019. "Decomposing treatment effect variation." *Journal of the American Statistical Association*, 114(525): 304–317.

Djebbari, Habiba, and Jeffrey Smith. 2008. "Heterogeneous impacts in PROGRESA." *Journal of Econometrics*, 145(1-2): 64–80.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition methods in economics." In *Handbook of labor economics*. Vol. 4, 1–102. Elsevier.

Goldsmith-Pinkham, Paul, Maxim Pinkovskiy, and Jacob Wallace. 2020. "Medicare and the Geography of Financial Health." *FRB of New York Staff Report*, , (911).

Heckman, James J, Jeffrey Smith, and Nancy Clements. 1997. "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts." *The Review of Economic Studies*, 64(4): 487–535.

Holland, Paul W. 1986. "Statistics and causal inference." *Journal of the American statistical Association*, 81(396): 945–960.

Kitagawa, Evelyn M. 1955. "Components of a difference between two rates." *Journal of the american statistical association*, 50(272): 1168–1194.

Kline, Patrick. 2011. "Oaxaca-Blinder as a reweighting estimator." *American Economic Review*, 101(3): 532–37.

Oaxaca, Ronald. 1973. "Male-female wage differentials in urban labor markets." *International economic review*, 693–709.

Słoczyński, Tymon. 2015. "The Oaxaca–Blinder unexplained component as a treatment effects estimator." *Oxford Bulletin of Economics and Statistics*, 77(4): 588–604.

United States Department of Health and Human Services. Centers for Disease Control and Prevention. 2020. "Behavioral Risk Factor Surveillance System (BRFSS) Survey Data, 2010-2018."

Wager, Stefan, and Susan Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

Wallace, Jacob, Karen Jiang, Paul Goldsmith-Pinkham, and Zirui Song. 2021. "Changes in Racial and Ethnic Disparities in Access to Care and Health Among US Adults at Age 65 Years." *JAMA Internal Medicine*, 181(9): 1207–1215.