Adjusting Risk Adjustment — Accounting for Variation in Diagnostic Intensity

Amy Finkelstein, Ph.D., Matthew Gentzkow, Ph.D., Peter Hull, B.A., and Heidi Williams, Ph.D.

In the U.S. health care system, payments and performance measures are often adjusted to account for differences in patients' baseline health and demographic characteristics. The idea behind such risk adjustments is to create a level playing field, so that providers aren't penalized for serving sicker or harder-to-treat patients and insurers aren't penalized for covering them. For example, the private insurance companies that participate in Medicare Advantage and the Affordable Care Act (ACA) exchanges receive risk-adjusted payments from the U.S. government, with the rationale that insurers should be reimbursed more for enrollees with higher expected costs.

The intent of risk adjustment is straightforward; implementing it in practice is far less so. The risk scores that are currently used are based on past diagnoses found in insurance-claims data, along with basic enrollee demographics. However, research has shown that differences in Medicare enrollees' reported diagnoses reflect not only differences in their underlying health (as intended) but also differences in their providers' proclivity for making diagnoses and recording them (which we refer to as "diagnostic intensity").[1] As a result, patients who live in areas of the country where providers tend to diagnose more aggressively will be measured as being sicker than patients who are actually similarly healthy but live in areas with less intensive diagnostic practices. Payments and performance measures will therefore be tilted to favor providers in more diagnostic-intensive regions.

These facts have important implications for health care policy. For example, variation in diagnostic intensity across providers that leads to bias in standard risk-adjustment measures could influence the results of comparative effectiveness studies. It may also distort risk-adjusted capitation and bundled payments for episodes of care, as well as risk-adjusted payments to private insurance companies participating in Medicare Advantage or the ACA exchanges.

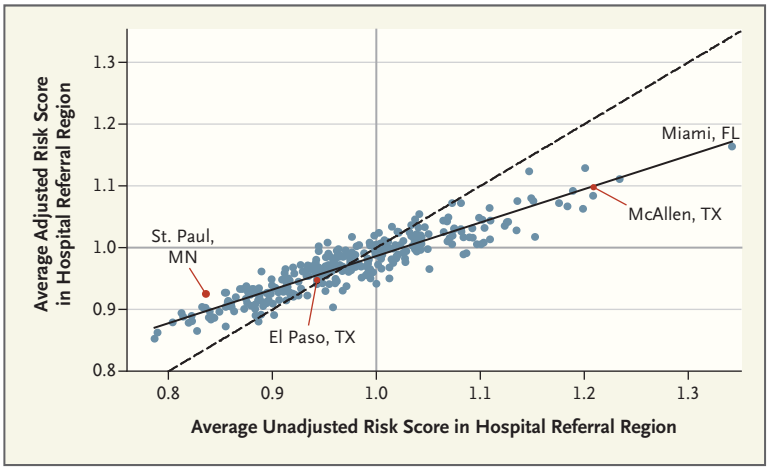
With place-specific risk-adjustment factors, health-based payments and performance measures could be scaled to counteract regional differences in diagnostic intensity.

We provide a concrete solution to this problem: place-specific risk-adjustment factors by which health-based payments and performance measures could be scaled to counteract regional differences in diagnostic intensity. To develop such an approach, we examined changes in the measured health of Medicare beneficiaries who move between different areas of the country. Prior work has documented that after Medicare beneficiaries move to a region where providers practice more intensely, they are measured as having more diseases.[1] We extended this basic insight of examining short-term changes in measured health in a narrow window around the time a patient moves; our analysis revealed that 50% of the geographic variation in measured health is due to place-specific differences in diagnostic practices[2] and yields estimates of place-specific adjustment factors (see the Supplementary Appendix, available at NEJM.org). These adjustment factors can be used by researchers or policymakers to correct regional estimates of population health or provider performance for place-specific measurement differences.

For illustrative purposes, we focused on risk scores used by the Centers for Medicare and Medicaid Services to adjust payments for Medicare Advantage plans. A risk score of 1 means the enrollee is predicted to have average Medicare spending relative to traditional Medicare enrollees, whereas a risk score of 1.1, for example, would indicate expected costs that are 10% above average. For each of the 306 hospital referral regions (HRRs) in the United States, we determined

the average risk score of traditional Medicare enrollees in that region, our estimated adjustment factor for that region, and the adjusted average risk score (see the Supplementary Appendix for a table listing these results for each region, which is also available as an Excel file).

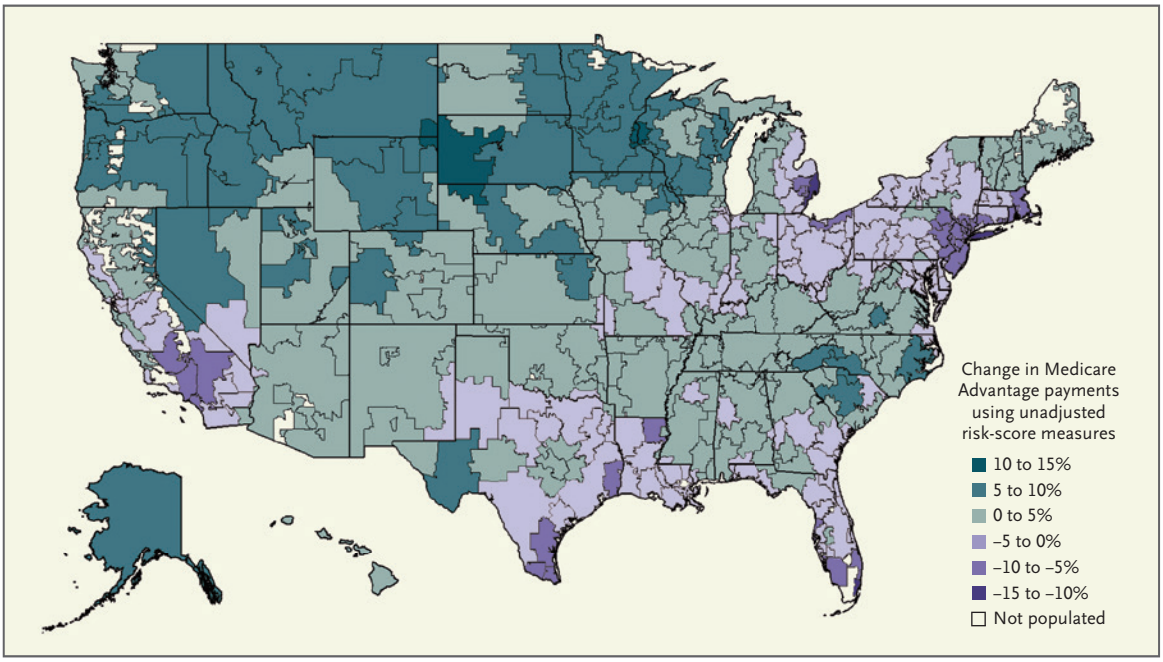
The graph shows the unadjusted average risk score and the adjusted risk score — which accounts for variation in diagnostic practices — for each HRR. The dashed line is a reference for the case in which our adjustment had no effect. The solid line is a linear regression line, showing the relationship between adjusted and unadjusted scores in practice.



Average Risk Scores for Traditional Medicare Patients by Hospital Referral Region.

Each point represents a hospital referral region. The dashed line represents equal adjusted and unadjusted average risk scores. The solid regression line reflects the relationship between adjusted and unadjusted scores in practice.

We found that actual patient health (the y axis) varies less across HRRs than one would assume on the basis of measured patient health (the x axis). Our analysis suggests that regions with enrollees whose actual health is worse tend to be regions where clinicians engage in more intensive diagnostic practices, so that measured differences in patient health are larger than the actual differences. The largest negative adjustment (13%) was for the Miami, Florida, region, and the largest positive adjustment (10%) was for the St. Paul, Minnesota, region.



Geographic Patterns in Risk-Score Adjustment.

Teal areas are regions where patients are less healthy than would be expected on the basis of raw risk scores; purple areas are regions where they are more healthy than would be expected.

The graph also highlights two HRRs in Texas that have been frequently discussed since the publication of Atul Gawande's *New Yorker* article "The Cost Conundrum": McAllen and El Paso.[3] Unadjusted risk scores suggest that patients in McAllen are 25% sicker than patients in El Paso, but our adjusted estimates suggest that patients in McAllen are only 15% sicker. In other words, in keeping with Gawande's conjecture, a nontrivial share of the difference in measured health between the two areas can be explained by more intensive diagnostic practices by doctors in McAllen than doctors in El Paso. If risk scores were adjusted to account for this finding, the gap in risk score–based reimbursements between McAllen and El Paso would shrink by about 10 percentage points.

The map shows the geographic patterns that emerge from our risk-score adjustment. Teal areas (in general, the West and Midwest) have patient populations that are less healthy than one would expect on the basis of raw risk scores — that is, they have providers who tend to diagnose less aggressively. Purple areas (in general, the Northeast and South) have patient populations that are more healthy than standard risk-adjustment measures would predict — that is, they have providers who tend to have greater diagnostic intensity. Medicare Advantage payments based on our adjusted risk score measures would tend to increase by as much as 10 to 15% in the dark teal areas of the map and fall by 10 to 15% in the dark purple areas.

Our goal here is to provide an example of how the area-specific adjustment factors we calculated may be useful in practice. We expect that the adjustment factors may also be useful in observational studies that compare the outcomes of patients in supposedly similar health who are exposed to different treatments, as well as in public reporting programs in which patient choices or physician referrals may be based on risk-adjusted quality measures.

Disclosure forms provided by the authors are available at NEJM.org.

From the Department of Economics, Massachusetts Institute of Technology (A.F., P.H., H.W.), and the National Bureau of Economic Research (A.F., M.G., H.W.) — both in Cambridge, MA; and the Department of Economics, Stanford University, Stanford, CA (M.G.).

1. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. N Engl J Med 2010; 363:45-53.
2. Finkelstein A, Gentzkow M, Williams H. Sources of geographic variation in health care: evidence from patient migration. Q J Econ 2016;131:1681-726.
3. Gawande A. The cost conundrum: what a Texas town can teach us about health care. The New Yorker. June 1, 2009.

DOI: 10.1056/NEJMp1613238

New Vaccines against Epidemic Infectious Diseases

John-Arne Røttingen, M.D., Ph.D., Dimitrios Gouglas, M.Sc., Mark Feinberg, M.D., Ph.D., Stanley Plotkin, M.D., Krishnaswamy V. Raghavan, Ph.D., Andrew Witty, B.A., Ruxandra Draghia-Akli, M.D., Ph.D., Paul Stoffels, M.D., and Peter Piot, M.D., Ph.D.

The vaccine-development response to the 2014 Ebola epidemic in West Africa, though a valiant effort, was too little, too late. Three vaccine candidates were tested successfully under challenging conditions.[1-3] Governments and foundations mobilized funds quickly. Companies and research-and-development institutions brought vaccine candidates into the field. Collaborations among the World Health Organization (WHO), funders, academia, civil society, and industry saw vaccines advancing through more than 15 accelerated clinical trials in a year. But the testing of Ebola vaccine candidates had previously stalled, though several candidates could have been ready for efficacy testing before the epidemic if the necessary investments had been made. In the absence of data on safety, immunogenicity, and dosing in humans, it was challenging to progress quickly with efficacy trials in West Africa. As a result, people who could have been protected instead became infected, and too many of them died. Moreover, there is no guarantee of similar risk-taking efforts in the future, especially given the poor market potential and the great clinical and regulatory uncertainties.

Vaccines can prevent outbreaks of emerging infectious disease from becoming humanitarian crises. The WHO recently deemed 11 pathogens as the most likely to cause severe outbreaks in the near future and will regularly update its list (see table). There are