1-1-2019

Targeting The Inequality In Quality Improvement

Raquel Maria Davis
sportchica17@gmail.com

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
Davis, Raquel Maria, "Targeting The Inequality In Quality Improvement" (2019). Public Health Theses. 1822.
https://elischolar.library.yale.edu/ysphtdl/1822

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Targeting the Inequality in Quality Improvement

Raquel Davis

Year Completed: 2019
Year Degree Awarded: 2019
Degree Awarded: Master of Public Health
Department: School of Public Health
Advisor: Dr. Mark Schlesinger
Committee Member: Dr. Amy Davidoff
Abstract

Objective. To use Casalino et al.’s (2007) proposed causal pathways of how pay-for-performance (P4P) can exacerbate racial/ethnic health care disparities as a framework for analyzing the literature and determine the adequacy of the evidence.

Data Sources/Study Setting. Existing literature (gray and peer-reviewed) on pay-for-performance and racial/ethnic health disparities identified from the Ovid Medline database.

Conceptual Framework. Casalino et al.’s four casual pathways include: “Pathway 1: Reduction in income for physicians in poor minority communities,” “Pathway 2: ‘Color-blind’ quality improvement programs,” “Pathway 3: ‘Teaching to the test’ might disproportionately affect minorities,” and “Pathway 4: Avoiding patients perceived as likely to lower quality scores.”

Principal Findings. For Pathway 1, there is consistent and strong evidence of harm. With a reduction in income for physicians in poor minority community settings, communities and their patients are being harmed. For Pathway 2, there is a lack of evidence. However, despite this, statistics from outside the P4P literature are consistent in their findings that “color-blind” approaches are occurring in health care. For Pathway 3, there is consistent and strong evidence of neglect. “Teaching to the test” is occurring, and existing P4P programs do not incorporate metrics that can improve disparities. Lastly, for Pathway 4, there is qualitative and quantitative evidence, but it is inconclusive. Some studies report avoidance whereas others do not.

Conclusion. Our health care system must step up to the challenge of accounting for differences in patient populations in payment models, such as P4P programs. If we do not, racial and ethnic minority patients will continue to suffer as structural inequities built into these programs persist.

Key Words. Pay-for-performance, racial and ethnic health care disparities, health policy
Acknowledgments

First, I would like to acknowledge Mark, who truly inspires students to make a difference. I would also like to acknowledge Amy, Quintin, and Lilo.

Table of Contents

A Pay-for-Performance Overview..............................................................p. 6

Pay-for-Performance Efficacy.................................................................p. 6

Racial and Ethnic Health Disparities and Pay-for-Performance.....................p. 8

Racial and Ethnic Disparities in Health Care Quality.................................p. 8

The Potential Effects of Pay-for-Performance on Health Disparities..............p. 9

Methods................................................................................................p. 10

The “Four Causal Pathways” of How Pay-for-Performance Can Exacerbate Health Disparities.................................................................p. 11

Pathway 1: Reduction in income for physicians in poor minority communities. ....p. 11
Pathway 2: “Color-blind” quality improvement programs. ..........................p. 14
Pathway 3: “Teaching to the test” might disproportionately affect minorities. ....p. 17
Pathway 4: Avoiding patients perceived as likely to lower quality scores. .........p. 20

Conclusion on the Pathways .................................................................p. 23

A Strategic Pay-for-Performance Policy......................................................p. 24

Discussion...............................................................................................p. 26

References...............................................................................................p. 28

Appendix.................................................................................................p. 33
Despite many technical resources and high medical spending, the quality of health care in the United States is not where it should be. To improve health care quality, payers are increasingly using financial incentives to reward physicians and medical groups that meet specific performance targets. Pay-for-performance (P4P), often referred to as value-based purchasing, is a specific payment model in which providers are rewarded through bonuses or penalized through reductions based on meeting preestablished targets or benchmarks for measures of quality and/or efficiency (Damberg et al., 2014). In theory, these initiatives should help drive the behavior of providers and health care systems to improve the quality of care delivered, reduce unnecessary use of expensive health care services, and improve patient health.
outcomes (Mendelson et al., 2017). However, even if P4P is effective in reaching these primary goals, whether P4P improves quality for all groups of patients needs to be considered.

Research has shown that the quality of health care for some groups of racial and ethnic minority patients (e.g. blacks and Hispanics) is worse than that for whites, and this leads to poorer health outcomes for these groups, and overall, lower life expectancy (Braveman et al., 2010; Gray et al., 2018; Smedley et al., 2003). Despite increasing attempts to develop interventions to help physicians and hospitals improve the quality of care for racial/ethnic minority patients and reduce the disparity that exists, the literature indicates that P4P programs are not explicitly designed to address racial/ethnic disparities (Chien et al., 2007; Rosenthal et al., 2007). Therefore, while P4P has the potential to reduce racial/ethnic disparities, it also has the potential to exacerbate them. Examining how this can occur is critical in mitigating these negative consequences.

This is where Casalino et al.’s article, “Will Pay-For-Performance And Quality Reporting Affect Health Disparities?” comes into play. Written in 2007, Casalino identifies four potential ways in which external incentives for quality may have the unintended consequence of increasing racial/ethnic health care disparities. Over 10 years have passed since the article was published, however. Therefore, my thesis will utilize Casalino’s proposed causal pathways as a framework for analyzing the literature and will add to the current body of knowledge surrounding P4P and racial/ethnic disparities. Furthermore, making use of the insights from the conceptual analysis, I will suggest ways of better deploying and researching P4P to reduce racial/ethnic disparities. Health care disparities remain one of our country’s major health challenges, and it is crucial that policymakers design programs and policies to minimize unintended consequences and reduce disparities.
A Pay-for-Performance Overview

The goal of pay-for-performance is to incentivize value by rewarding higher quality care at lower cost, and the underlying P4P assumption is that improvements in quality are under the control of the provider (Damberg et al., 2014). Other factors that influence the success of P4P programs include the external environment, practice characteristics, and program features (Chee et al., 2016). The external environment includes regulations, policies, the patient population, patient preferences, and other quality improvement initiatives which can either promote or prevent the potential success of P4P programs (Chee et al., 2016). Practice characteristics include structure of the health care system, organizational culture, and available resources and capabilities (especially in information technology) (Chee et al., 2016). Program features include defining the patient population, the program goals, measures, financial incentive, and risk structure (Chee et al., 2016). In addition to these factors, key elements that must be carefully considered when designing and implementing P4P programs include provider engagement, individual versus group motivators, sizeable incentives, measure alignment and selecting high-impact performance measures, making payment reward all high-quality care, data and other quality improvement support, and prioritizing quality improvement for underserved populations (Damberg et al., 2014; Rosenthal & Dudley, 2007).

Pay-for-Performance Efficacy

In 2014, Damberg and colleagues with RAND were asked by the U.S. Department of Health and Human Services to evaluate available literature on Value-Based Purchasing models, including pay-for-performance (Damberg et al., 2014). The researchers identified 49 studies that examined the effect of P4P on process and intermediate outcome measures (37 studies examined
the effect of P4P on process measures for physicians or physician groups, 11 studies examined the effect of P4P on process measures in the hospital setting, and a single study examined the effect of P4P on process measures in other care settings) and 21 studies that evaluated the effect of P4P on outcomes (12 outcomes studies in physician groups, 6 in hospitals, and 3 in other settings) (Damberg et al., 2014). Overall, the results of the studies were mixed, and studies with stronger methodological designs were less likely to identify significant improvements associated with the P4P programs (Damberg et al., 2014). Any identified effects were relatively small (Damberg et al., 2014).

With the literature having grown considerably since the RAND review, Mendelson and colleagues conducted another systematic review in 2017 with the funding support of the U.S. Department of Veterans Affairs Quality Enhancement Research Initiative (Mendelson et al., 2017). This review ultimately included 69 studies (58 were in ambulatory settings, 52 reported process-of-care outcomes, and 38 reported patient outcomes) (Mendelson et al., 2017). In the ambulatory setting, the review found low-strength evidence that P4P programs may improve process-of-care outcomes over the short term (2 to 3 years), but evidence on the longer-term effects was limited (Mendelson et al., 2017). Many of the studies reporting positive findings were conducted in the United Kingdom, where incentives are much larger than any P4P program in the United States (Mendelson et al., 2017). The review found low-strength evidence that P4P had little to no effect on intermediate health outcomes (e.g. changes in laboratory measures), despite inconsistencies among study results (Mendelson et al., 2017). Lastly, the evidence examining patient health outcomes was insufficient, because few methodologically rigorous studies reported these outcomes (Mendelson et al., 2017). In summary, the researchers found
low-strength, contradictory evidence that P4P programs could improve processes of care but found no clear evidence suggesting that they improve patient outcomes (Mendelson et al., 2017).

Overall, the published evidence regarding improvements in performance from the P4P experiments of the past decade is mixed, and where observed, improvements were typically modest. Many of the published studies evaluating the impact of P4P programs suffer from methodological weaknesses that make it hard to determine whether the P4P intervention had an effect above and beyond other changes (e.g., investment in quality improvement support, public reporting, health information technology investments and support) that were simultaneously occurring to improve quality and restrain spending (Damberg et al., 2014; Mendelson et al., 2017). More research is needed in order to generate the information required to fill gaps in the P4P knowledge base.

Racial and Ethnic Health Disparities and Pay-for-Performance

Racial and Ethnic Disparities in Health Care Quality

Although equity is one of the six dimensions of quality defined by the Institute of Medicine, well-documented disparities in care between racial/ethnic minority patients and non-Hispanic white patients persist (Institute of Medicine (US) Committee on Quality of Health Care in America, 2001; Smedley et al., 2003). Figures 1 (Figure 19) and 2 (Figure 20) in the Appendix, directly from the Agency for Healthcare Research and Quality’s “2017 National Healthcare Quality and Disparities Report,” demonstrate these quality differences (Gray et al., 2018). With decreased quality of health care, some groups of racial/ethnic minority patients (e.g. blacks and Hispanics) have poorer health outcomes and overall, lower life expectancy compared to whites (Braveman et al., 2010; Gray et al. 2018; Smedley et al., 2003).
Of importance to note is that minority patients vary tremendously in income, education, acculturation to the United States, and English language ability (Casalino, Elster, et al., 2007). Unintended consequences of the health care system are more severe for the more disadvantaged range of this spectrum (e.g. low-income blacks) (Braveman et al., 2010). For instance, Braveman et al. found that for many child and adult indicators examined (e.g. infant mortality, life expectancy at age 25, and self-reported health status), socioeconomic differences within black and white racial/ethnic groups were at least as striking as socioeconomic differences overall (Braveman et al., 2010). Because of these differences, this thesis will have a particular focus on the most disadvantaged racial/ethnic groups, who consistently experience worse quality of care than whites. This will become apparent in the analysis section, as low-income black, Hispanic, and American Indians/Alaska Native patients are referred to most frequently. Therefore, the working definition of “disparities” in this report will reflect overall differences in quality of care and outcomes, as opposed to after adjustment for socioeconomic status, underlying health status, etc. Furthermore, the working definition of “quality” will be broadly defined in that it encompasses structure, process, and outcome measures.

The Potential Effects of Pay-for Performance on Health Disparities

Increasing attempts have been made to develop interventions to help physicians and hospitals improve the quality of care for racial and ethnic minority patients and reduce the disparity that exists (Smedley et al., 2003). It has been acknowledged that one such intervention could be P4P programs. The Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care stated in its report, “Economic incentives should be considered for practices that improve provider-patient communication and trust, and reward appropriate screening, preventive, and evidence-based clinical care. In addition, payment systems should
avoid fragmentation of health plans along socioeconomic lines” (Smedley et al., 2003). As mentioned, however, health systems must be careful in implementing and evaluating these policies, as P4P incentives have not been explicitly designed to address racial/ethnic disparities and could have the outcomes of exacerbating disparities, improving quality for all groups equally, or even attenuating disparities (Chien et al., 2007; Rosenthal et al., 2007; Shakir et al., 2018). Minute details of the measures, formulas for calculating performance, and rules for defining rewards and penalties can have substantial impact on “who wins and who loses,” and on unintended consequences, such as potentially worsening health care disparities (Chee et al., 2016).

Evidence to support substantive conclusions regarding the largest and most recent P4P programs and their effects on disparities is mixed and still insufficient (Shakir et al., 2018). The evidence we do have suggests that these programs at least have the potential to reduce disparities (Bhalla et al., 2013; Petersen et al., 2017; Song et al., 2017), but their effects vary widely, and further investigation is necessary (Shakir et al., 2018). Furthermore, examinations of whether P4P programs work to reduce or increase disparities are challenged by the lack of information at the patient level on race, ethnicity, education, SES, and other markers of vulnerable populations prone to disparities (Damberg et al., 2014).

Methods

Articles and studies were identified from the Ovid Medline database. A fairly extensive literature review was conducted of the pay-for-performance and racial/ethnic health disparity literature (gray and peer-reviewed). However, when extrapolation outside the P4P literature was needed, the review of the literature was to a lesser degree.
The “Four Causal Pathways” of How Pay-for-Performance Can Exacerbate Health Disparities

At the time Casalino et al.’s 2007 paper was written, there was limited data on the effects of performance incentives on racial and ethnic disparities in quality (Casalino, Elster, et al., 2007). Casalino et al.’s paper is the first of its kind to suggest causal pathways, four specifically, in which P4P programs (and public-quality reporting programs) have the potential to increase disparities in health care delivery. (A fifth causal pathway is given, but it only applies to public reporting.) Specifically, Casalino et al. focus on physicians but state that analogous points can be made for hospitals and health plans (Casalino, Elster, et al., 2007).

The objective of my paper is to use the current literature in analyzing these pathways and determine the evidentiary strength for each one. What should be recognized is that these pathways could work differently for varying P4P intensities (e.g. 5 percent of provider income vs. 25 percent of provider income), provider levels (e.g. solo practice vs. a large hospital setting), and patient characteristics (e.g. minority patients with one vs. multiple chronic diseases). However, due to limited P4P and racial/ethnic minority literature, I take a broader approach in this analysis, such as how Casalino et al. did.

Pathway 1: Reduction in income for physicians in poor minority communities.

Casalino et al. hypothesized that P4P may adversely affect the income of physicians practicing in minority communities, particularly poor minority communities, and subsequently, reduce both the number of physicians working in such communities and their ability to invest in processes to improve quality (Casalino, Elster, et al., 2007). At the time Casalino et al.’s work was published, there was only one prior research study that suggested evidence for this “poor may get poorer” and “rich may get richer” phenomenon (Casalino, Elster, et al., 2007). This first
evaluation of a P4P program introduced for primary care physicians by the British National Health Service (NHS) found that practices serving lower-income populations did, in fact, have lower P4P quality scores and subsequently, received less earnings (Doran et al., 2006). Since Casalino’s article, more research has been published, and there is more clarity surrounding this issue.

In 2008, Werner et al. published a longitudinal study of the relationship between hospital quality performance and percentage Medicaid coverage from 2004 to 2006 (Werner, Goldman, & Dudley, 2008). This study found that hospitals with high percentages of Medicaid patients had worse performance in 2004 and had significantly smaller improvements over time than those with low percentages of Medicaid patients (Werner et al., 2008). For instance, hospitals with low percentages of Medicaid patients improved composite acute myocardial infarction performance by 3.8 percentage points vs. 2.3 percentage points for those with high percentages, an absolute difference of 1.5 (p=.03) (Werner et al., 2008). This resulted in a relative difference in performance gains of 39 percent (Werner et al., 2008). The article concluded that incentives based on measures being used did have the potential to increase disparities among hospitals (Werner et al., 2008). Another hospital-based study with similar findings was published in 2013 by Ryan. Ryan examined incentive payments in the first year of Medicare’s Hospital-Value Based Purchasing (HVBP) program (The program began in October of 2012.) to determine whether hospitals that care for more patients who are disadvantaged received lower payments (Ryan, 2013). Using a Disproportionate Share Hospital (DSH) index, with a higher DSH indicating the hospital’s patient population is at a greater socioeconomic disadvantage, Ryan found that hospitals with a higher DSH index value had significantly lower Medicare payment adjustments (p < 0.01) in the first year of HVBP, which resulted in a significantly more negative
expected financial impact (Ryan, 2013). For instance, hospitals at the 5th percentile of the DSH index expect a financial impact of +$18,900 (95% confidence interval), whereas those at the 95th percentile expect an impact of -$23,300 (95% confidence interval) (Ryan, 2013). These results demonstrate that hospitals caring for more disadvantaged patients did, in fact, fare worse in the first year of HVBP.

In the primary care setting, Friedberg et al. published a study in 2010 in which they simulated P4P payments to 438 Massachusetts practices serving higher and lower shares of patients from vulnerable communities (Friedberg et al., 2010). Various preventive care categories were examined, ranging from cancer to diabetes screening, for practices with low to high racial/ethnic vulnerability and economic vulnerability (Friedberg et al., 2010). The authors found no instance in which practices serving the most vulnerable patient catchment areas would receive higher payments than other practices (Friedberg et al., 2010). In fact, through their simulation, they found that whereas the practice with low racial/ethnic vulnerability would receive a $15 median annual payment per patient for preventive care, the practice with high vulnerability would only receive $9 (p < .01) (Friedberg et al., 2010). Furthermore, total payment per practice would be $9,300 for low racial/ethnic vulnerability compared to $6,200 for high (p < .01) and per physician payment in practice would be $2,300 for low racial/ethnic vulnerability compared to $1,600 for high (p < .01) (Friedberg et al., 2010). These findings demonstrate that practices serving larger proportions of vulnerable populations would receive lower simulated payments for preventive care than other practices.

The empirical literature overwhelmingly supports Pathway 1. Consistently documented and all-around statistically significant large effects (p < .05) indicate a reduction in income for physicians in poor minority community settings due to P4P incentives, and as a result,
communities and their patients are being harmed. The questions that then must be asked include: How much of this variation is internal to the practice versus external? Is this a practice-level problem or a social-determinant-of-health-level problem? In answering these and attempting to determine the causality of this issue, a policy approach can be better appropriately targeted.

**Pathway 2: “Color-blind” quality improvement programs.**

Casalino et al. claimed that P4P could increase disparities if it induces medical groups to create organized processes to improve quality, such as programs to help patients with chronic illnesses learn “self-management” of their illness, and if minority patients are less likely to benefit from these programs (Casalino, Elster, et al., 2007). An example Casalino provides is if education is given in a language or cultural style or at a literacy level that minority patients understand poorly or feel is culturally insensitive (Casalino, Elster, et al., 2007). The literature was sparse surrounding this proposed pathway when Casalino referenced it and is still sparse. In presenting this pathway, Casalino did not reference a particular P4P study but rather, cited general clinical studies that indicated a need to make approaches tailored (Lindau et al., 2006; Pham et al., 2005). Therefore, like Casalino does, inferences can be made in attempting to draw conclusions.

While literature is scarce surrounding the potential impact of P4P on “color-blind” quality improvement programs, other literature indicates that quality improvement is, to at least some extent, “color-blind.” A 1996 empirical study linked 1990 census data on median income according to ZIP Code with 1993 Medicare administrative data for 26.3 million beneficiaries 65 years of age or older (24.2 million whites and 2.1 million blacks) (Gornick et al., 1996). The authors calculated age- and sex-adjusted rates of various diagnoses and procedures according to race and income and computed black to white ratios (Gornick et al., 1996). The study found that
for every 100 women, there were 26.0 mammograms among whites and 17.1 mammograms among blacks (black:white ratio of .66, p < .001), and for every 1,000 beneficiaries, there were 515 influenza immunizations among whites and 313 among blacks (a black:white ratio of .61, p < .001) (Gornick et al., 1996). The study concluded that race had substantial effects on use of services among Medicare beneficiaries and that Medicare coverage alone is not sufficient to promote effective patterns of preventive health service use by all beneficiaries (Gornick et al., 1996). A little more recently, findings reported by the Commonwealth Fund’s 2006 “Racial and Ethnic Disparities in U.S. Health Care: A Chartbook,” which used a variety of sources ranging from national surveys to single-site studies, demonstrated differences in experiences of services (Mead et al., 2008). For instance, black women were more likely to receive inadequate communication of their mammography screening results compared with white women, especially if the results were abnormal (Mead et al., 2008). Furthermore, black breast cancer patients were less likely to receive a complete diagnostic evaluation within 30 days of a patient-noted abnormality or abnormal mammogram (Mead et al., 2008). Hispanics were less likely to adhere to antiretroviral therapy, in part due to language barriers and lack of interpreters (Mead et al., 2008). Blacks and American Indians/Alaska Natives were more likely than whites to have low birthweight babies, which could be diminished with timely prenatal care (Mead et al., 2008). Despite effect size not being recorded in this report, these are some of many examples in which participation in prevention efforts by minority patients is less than that of whites and where medical groups are not appropriately targeting quality prevention services and programs to meet the needs of racial/ethnic minority patients.

A later 2007 paper in which Casalino was involved further attempted to examine this pathway by identifying two empirical studies that would suggest a “color-blind” approach in a
P4P program could result in maintaining disparity levels (Chien et al., 2007). One longitudinal study of 58,700 hemodialysis patients from 1993 through 2000 used a one-size-fits-all approach for improving hemodialysis dosing for patients with end-stage renal disease, without the recognition that disease burden and treatment choices may differ across race and ethnicity (Sehgal, 2003). It led to an overall 40 percent improvement in hemodialysis dosing across the 8-year study period, but did not change the disparity between black and white patients in anemia management and nutritional status (Sehgal, 2003). In contrast, a multisite randomized clinical trial of 1,801 older adults comparing collaborative care for depression with treatment as usual in primary care had an inclusive approach by recognizing that disease burden and treatment choices were different across race and ethnicity improved depression care (Areán et al., 2005). The use of antidepressants by minority and nonminority elderly patients improved and the gap between minorities and nonminorities closed, even though minorities had more severe disease and less antidepressant use at the beginning of the intervention (Areán et al., 2005).

Furthering this point, an example was given in the literature in which structural quality measures were specifically aimed at cultural competence. For some time, Massachusetts required hospitals to report on a range of activities related to caring for minority patients by using a checklist known as the Cultural Competence Organizational Self-Assessment (Blustein et al., 2011). Sample items included “board of directors and senior management reflect the racial/ethnic mix of the actual population mix being served,” “policies exist to include racial/ethnic communities in the planning and design of health care services,” and “patient surveys are translated for non-English-speaking patients” (Blustein et al., 2011). For each of these items, hospitals were required to indicate their current level of activity (Blustein et al., 2011). In 2010, Massachusetts decided it would no longer use these structural measures due to some items being
unclear and subjective (Blustein et al., 2011). However, nationwide, the interest in cultural competence, and possibly inclusive quality improvement programs, is continuing (Blustein et al., 2011).

Concluding this pathway, there are no studies specifically analyzing how P4P induces medical groups to create organized processes to improve quality that have less benefits for minority patients. However, despite this lack of evidence, statistics from outside the P4P literature are consistent in their findings that “color-blind” approaches are occurring in health care. As a result, minority patients are using less preventive services partly due to their specific needs not being met. These findings indicate that organized processes to improve quality must factor cultural competence into their designs. Massachusetts was an example in which hospitals were being rewarded for making approaches more tailored. When analyzing this pathway, what must be considered is that equality (e.g. whites and minorities both improving by X percent and a disparity gap still existing) does not equate to equity (e.g. decreasing the disparity gap by recognizing that minority groups will need specialized and likely more services). Therefore, in order to improve disparities, P4P programs must acknowledge differences and place emphasis on creating tailored approaches to minority populations.

Pathway 3: “Teaching to the test” might disproportionately affect minorities.

P4P programs tend to identify multiple high-level goals that focus on improving clinical quality (roughly 75 percent of programs) and cost/affordability (roughly 53 percent of programs) (Damberg et al., 2014). However, less common are goals related to improving patient outcomes (34 percent of programs) and patient experience (17 percent of programs) (Damberg et al., 2014). Casalino et al. claimed that these selective P4P goals and subsequently, measurements could induce physicians to focus their time and attention (consciously or unconsciously) on types
of care that are being measured, to the detriment of non-measured areas that could be equally or more important (Casalino, Elster, et al., 2007). Casalino et al. believed this “teaching to the test,” a phenomenon which has also been found in other industries, could disproportionately affect minorities (Casalino, Elster, et al., 2007). For example, efforts targeting certain diseases, such as HIV and mental health disorders, may be more relevant to vulnerable patients than metrics such as hospital-acquired infections, time to acute myocardial infarction treatment, or readmission rates (Shakir et al., 2018).

In another 2007 article, Casalino et al. conducted a national survey of 556 general internists’ views on pay-for-performance and public reporting of quality scores (Casalino, Alexander, et al., 2007). When asked if measuring quality would divert physicians’ attention from important types of care for which quality is not measured, 22 percent of general internists strongly agreed with this statement, 39 percent somewhat agreed, 29 percent disagreed, and 11 percent strongly disagreed (Casalino, Alexander, et al., 2007). Therefore, a majority believed that “teaching to the test” would, in fact, occur (Casalino, Alexander, et al., 2007).

The quantitative literature also suggests this “teaching to the test” phenomenon occurs. A paper by Schlesinger et al. highlights this with two United Kingdom studies (Schlesinger et al., 2015). One study specifically investigated whether the UK’s incentive scheme, the Quality and Outcomes Framework (QOF), led general practitioners to neglect activities not included in the scheme (Doran et al., 2011). Through conducting a longitudinal analysis of achievement rates for 42 activities (23 included in the incentive scheme and 19 not included), the authors found that aspects of quality that were incentivized did better than predicted from previous trends, whereas those not incentivized fell below their trend lines (Doran et al., 2011). A second study of the
UK’s QOF found that because patient experience was only 2 of 146 incentivized metrics, patient valued aspects of quality did not have success (Campbell et al., 2010).

These qualitative and quantitative research findings indicate that physicians are focusing their time and attention on types of care that are being measured but neglecting care that is not being measured. The question that then must be asked is, “Are we even measuring the right things?” If we are not incentivizing metrics that are relevant and important to racial and minority populations, then P4P will, in fact, exacerbate health disparities.

In 2017, Eggleton et al. developed a taxonomy of measures used in the Merit-Based Payment System (MIPS) to assess what measurement gaps exist if a more specific focus on health disparities is to occur (Eggleton, Liaw, & Bazemore, 2017). By applying a theoretical model of health care quality as a coding matrix, the authors used a directed content analysis approach to categorize individual MIPS measures (Eggleton et al., 2017). In their analysis, the authors found that most MIPS measures related to aspects of clinical effectiveness, whereas few, if any, related to aspects of patient experience or interpersonal care (Eggleton et al., 2017). For instance, whereas an outcome measure such as “Experiences of patients in receiving health care” had one applicable measure (< 1 percent), a process measure such as “The effective application of knowledge-based care” had 122 (78 percent) applicable measures. These gaps suggest that MIPS, and other similar programs, may fail to measure the broader aspects of health care quality and even risk worsening existing disparities (Eggleton et al., 2017). However, if we are wanting to look at clinical outcomes, then current measures still do not address certain conditions that are of particular relevance to racial/ethnic minorities. For instance, HIV was previously mentioned as a disease that could be more relevant to vulnerable patients. Yet, looking at California’s Integrated Healthcare Association’s “Align. Measure. Perform. (AMP) Measure Set for
Measurement Year 2019,” one of the largest P4P programs, HIV is not mentioned in a single metric (Integrated Healthcare Association, 2019).

Both the qualitative and quantitative literature evidence is consistent and strong for Pathway 3. “Teaching to the test” is occurring. Furthermore, there are evident problems with the narrow set of measures typically being used in P4P programs (Damberg et al., 2014; Wharam et al., 2011). When attempting to select more appropriate measurements with a specific focus on improving health disparities, health systems must take into account two things. First, health systems must consider whether equity is being appropriately measured (Anderson et al., 2018; Hasnain-Wynia, 2007; Rubin, 2018). Second, health systems must ask themselves if they are measuring metrics important to racial/ethnic minority populations, such as doctor communication or HIV metrics (Collins et al., 2017; Shakir et al., 2018).

**Pathway 4: Avoiding patients perceived as likely to lower quality scores.**

Casalino et al. stated that P4P might induce physicians and medical groups to avoid patients whom they perceive as being likely to lower their quality scores, particularly if quality measures are not adequately adjusted for the patients’ overall health status and perhaps for racial or socioeconomic characteristics as well (Casalino, Elster, et al., 2007). In the literature, such occurrences have been referred to as “racial profiling” (Werner Rachel M., Asch David A., & Polsky Daniel, 2005), “exception reporting” (Doran et al., 2006), “cherry-picking” (Chien et al., 2007), and “risk selection” (Petersen et al., 2017). Casalino et al. found evidence that physicians perceive minority patients as less likely to comply with their recommendations for treatment and preventive services and more likely to have bad outcomes (Casalino, Elster, et al., 2007). They also found evidence that physicians subject to external incentives will try to avoid minority patients because they perceive them as more likely to have poor outcomes from treatments, even
when this is not the case (Casalino, Elster, et al., 2007). Since Casalino et al.’s paper, several other papers have discussed patient avoidance in the P4P setting.

In a 2009 qualitative study conducted by McDonald and Roland, 40 primary care physicians (20 from California and 20 from the UK) were interviewed about the unintended consequences of P4P (McDonald & Roland, 2009). A common theme that emerged included “Threats to the Ongoing Patient-Physician Relationship” (McDonald & Roland, 2009). Under this theme, it was reported that Californian physicians affiliated with the largest financial incentives expressed frustration about patients who refused to comply with their advice and the inability to exclude individual patients from performance calculations (McDonald & Roland, 2009). As a result, some physicians reported behaviors such as forced disenrollment of noncompliant patients and other strategies reported by physicians included accusing patients of damaging their physician’s rating or lying to patients about the financial consequences of their refusing to comply (McDonald & Roland, 2009). On the other hand, English physicians, who were allowed to exclude or file an “exception report,” did not report feeling resentment or resorting to the methods used by the Californian physicians (McDonald & Roland, 2009).

Whereas McDonald and Roland’s article gives us reason to believe patient avoidance is occurring, two quantitative studies have demonstrated otherwise. Using 2000 to 2006 Medicare data, a 2010 study by Ryan tested for differences in the conditional probability of receiving care at Premier Inc. Hospital Quality Incentive Demonstration (PHQID) hospitals, which utilized pay-for-performance (P4P) programs, for acute myocardial infarction (AMI), heart failure, and pneumonia before and after implementation of the PHQID between white and minority patients (Ryan, 2010). The author also tested for differences in the conditional probability that white and minority patients diagnosed with AMI received coronary artery bypass graft (CABG) in hospitals
participating and not participating in the PHQID before and after the implementation of the PHQID (Ryan, 2010). Ryan found little evidence that the PHQID reduced access for minority patients (Ryan, 2010). Only “Other Race” beneficiaries (It is unclear whether Other Race in this study consisted of unknown race categories or mixed race categories.) had a significant reduction (p < .05) in adjusted admissions to PHQID hospitals in the postperiod, and this was only for AMI (Ryan, 2010). Only marginally significant (p < .10) evidence of a reduction in CABG was found, and this also only occurred for Other Race beneficiaries (Ryan, 2010). Another quantitative study, published in 2016 by Petersen et al., evaluated the effect of a 2007 to 2009 P4P intervention on the quality of hypertension care provided by Veteran Affairs physicians and their primary care panels to black patients and whether it produced risk selection (Petersen et al., 2017). Using a nested study within a cluster randomized controlled trial of three types of financial incentives and no incentives (control), risk selection was measured by comparing the proportion of patients who switched providers, patient visit frequency, and panel turnover (Petersen et al., 2017). The authors found that this P4P intervention did not produce risk selection, as there was no difference between intervention and controls in the proportion of patients who switched providers, visit frequency, or panel turnover (Petersen et al., 2017).

The Ryan et al. and Petersen et al. studies examine racial/ethnic minority patients but do not account for additional factors that could result in avoidance, such as minority patients with and without (or perceived by clinicians to have) chronic conditions. In the UK, several P4P studies by Millet et al. have looked at addressing racial/ethnic disparities in the context of the management and control of chronic diseases, and one particular study specifically analyzed exclusion reporting. In this 2011 study, three cross-sectional analyses were conducted using data from the electronic medical records of all patients with diabetes registered in 23 general practices
in Brent, North West London between 2004 to 2005 and 2006 to 2007 (Dalton et al., 2011). Patterns of exclusions were examined for three intermediate outcome indicators (HbA1c, blood pressure, and cholesterol) (Dalton et al., 2011). Study results indicated that black and South Asian patients were more likely to be excluded from the HbA1c indicator than White patients [adjusted odds ratio = 1.64 (1.17–2.29) in 2005 to 2006] (Dalton et al., 2011). Furthermore, patients registered with practices in deprived areas were consistently exception reported at higher levels (Dalton et al., 2011). This finding mirrors those of Sigrid et al. in an ecological study of diabetes exception reporting (Dalton et al., 2011).

The literature for Pathway 4 is inconsistent. The qualitative literature suggests avoidance can occur. However, the quantitative literature has mixed evidence, with significant evidence both for and against avoidance occurring. In attempting to gain more clarity, I did turn to the public reporting literature, but mixed evidence was also prevalent in these studies as well. Because of lack of conclusive evidence for P4P and other incentive programs, monitoring of avoidance must continue (Ryan, 2010).

Conclusion on the Pathways

For “Pathway 1: Reduction in income for physicians in poor minority communities,” there is consistent and strong evidence of harm. With a reduction in income for physicians in poor minority community settings, communities and their patients are being harmed. For “Pathway 2: ‘Color-blind’ quality improvement programs,” there is a lack of evidence. There are no studies specifically analyzing how P4P induces medical groups to create organized processes to improve quality that have less benefits for minority patients. However, despite this lack of evidence, statistics from outside the P4P literature are consistent in their findings that “color-blind” approaches are occurring in health care, and a result, minority patients are using less
preventive services. For “Pathway 3: ‘Teaching to the test’ might disproportionately affect minorities,” there is consistent and strong evidence of neglect. “Teaching to the test” is occurring, and existing P4P programs do no incorporate metrics that can improve disparities. Lastly, for “Pathway 4: Avoiding patients perceived as likely to lower quality scores,” there is qualitative and quantitative evidence, but it is inconclusive. Some studies report avoidance whereas others do not. Reference Table 1 for a summary of the studies used in the analysis process.

A Strategic Pay-for-Performance Policy

A secondary objective of this paper is to consider a policy approach that targets one of Casalino’s pathways. What is important to note is that none of the pathways were “disproven” in the analysis, and all are viable options to target, either through further research or new policy efforts. However, because “Pathway 1: Reduction in income for physicians in poor minority communities” had consistent and strong evidence of harm, it would be easiest to insist that policymakers make use of these findings. “Pathway 3: ‘Teaching to the test’ might disproportionately affect minorities” also had consistent and strong evidence, but evidence demonstrated neglect, which is harder to “prove” to policymakers, unfortunately, over blatant harm.

Therefore, from Pathway 1, it can be inferred that minority patients with low SES and the providers who serve them in these safety-net settings appear to be some of the most vulnerable targets in regards to P4P exacerbating racial/ethnic health disparities. This supports the previous point in this paper that unintended health consequences are more severe for the more disadvantaged range of this spectrum and paying special attention to racial and socioeconomic
patterns in health jointly can inform policies to address inequalities in both dimensions (Braveman et al., 2010; Casalino, Elster, et al., 2007). Reference Figure 3, which displays race/ethnicities and percentage with coronary heart disease (CHD), to see an example of this gradient (Braveman et al., 2010).

In regards to a policy proposal, Werner, Ryan, and Friedberg’s studies on safety net providers in P4P programs being penalized financially suggest that currently used quality measures fail to account for differences in patient populations that could skew quality scores in favor of practices that care for higher-income, better-educated, and less-complex patients (Rubin, 2018). Measures do not adequately translate differences in care among groups into a metric or statistic that indicates the extent of disparities, a component very important when designing P4P programs and attempting to mitigate health disparities (Blustein et al., 2011). Casalino et al.’s 2007 general internist article validates this claim, with 85 percent of physicians believing that measures of quality are not adequately being adjusted for factors such as patients’ socioeconomic status (Rubin, 2018). Therefore, P4P programs must create a metric or statistic that adequately takes into account the patient population of a given provider or group of providers. In doing so, providers will not be penalized for external factors they cannot control, such as social-determinant-of-health-level issues.

Undertaking this type of policy would pose a challenge to many P4P programs. In order to differentiate between internal versus external factors affecting quality scores, data at the patient level needs to be collected on race/ethnicity and furthermore, socioeconomic and health status (Blustein et al., 2011). However, there is a lack of routinely collected data on demographic and socioeconomic characteristics of patients to begin with (Damberg et al., 2014). As stated in a 2018 JAMA article by a cardiologist and health services researcher, “We don’t have quality
measures that adequately sort out differences in quality vs. differences in patient population. We don’t even have the ability to collect and report high-quality clinical data…and that’s worse in under resourced settings” (Rubin, 2018). Furthermore, for underresourced practices with limited office time and resources, collecting such data will be difficult and imposing such a request on these practices could further burden the racial/ethnic minority patients whom they serve. Therefore, a policy which creates a metric or statistic to indicate the extent of disparities for a given office must provide underresourced practices with funding for the necessary personnel and infrastructure to collect this data.

Overall, the policy I am proposing will require all practices involved in a P4P program to collect necessary patient-level data, with assistance given to those practices that have struggled in past payment cycles. Then, once data is collected, payers can calculate a metric or statistic that then allows for risk adjustment of a practice’s total quality score. The most important aspect to consider as well is that the policy will be viewed positively by minority populations and will not result in further disparities.

Similar policies to this have begun to arise. In 2020, MIPS will begin adjusting payments based on the proportion of patients in a practice that are eligible for Medicaid, a marker for lower SES, as well as Medicare (Rubin, 2018). It has been indicated that this policy can be done on a rapid timeline, and though it may not be enough, policies such as this are a step in the right direction (Rubin, 2018).

Discussion

This thesis added to the pay-for-performance and racial/ethnic disparity literature. It used Casalino et al.’s causal pathways of how P4P can exacerbate racial/ethnic health care disparities, which were proposed over 10 years ago, as a framework and determined the adequacy of current
available evidence. This thesis highlighted that further research must be conducted and subsequently, policy change.

As previously mentioned, the causal pathways could work differently for varying P4P intensities (e.g. 5 percent of provider income vs. 25 percent of provider income), provider contexts (e.g. solo practice vs. a large hospital setting), and patient characteristics (e.g. minority patients with one vs. multiple chronic diseases). Therefore, a P4P/health disparity research agenda moving forward must recognize this, and studies must be specific in answering questions with these variances in mind. Reference Table 2 in the Appendix for examples of how future research should attempt to hone in on these differences, as this could further help with targeted policy approaches.

My thesis had a few limitations. First, Casalino et al.’s causal pathways were for both P4P and public reporting programs. By solely focusing on P4P programs, I only analyzed four of the five pathways Casalino had in mind. Second, due to the lack of literature on P4P and racial/ethnic disparities, my analysis had a broad rather than narrow focus and had to extrapolate from literature outside of P4P.

In conclusion, the design features and the context in which a P4P program is implemented are critical determinants of program success, especially when attempting to reduce health disparities through P4P (Damberg et al., 2014). Although the past decade has witnessed a fair amount of experimentation with performance-based-payment models, we still know very little about how best to design and implement them in order to achieve desired goals, the optimal conditions for their successful implementation, and provider response to the incentives (Damberg et al., 2014). Despite this uncertainty, our health care system must step up to the challenge of accounting for differences in patient populations in payment models, such as P4P programs. If
we do not, racial and ethnic minority patients will continue to suffer as structural inequities built into these programs persist.

References


Clinical Practice. Health services research, 50 Suppl 2(Suppl Suppl 2), 2116–2154.


### Appendix

**Table 1: Causal Pathway Analysis**

<table>
<thead>
<tr>
<th>Causal Pathways</th>
<th>Supporting Studies</th>
<th>Study Summaries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pathway 1: Reduction in income for physicians in poor minority communities</strong></td>
<td>Werner et al. (2008)</td>
<td>A longitudinal study of the relationship between hospital quality performance and percentage Medicaid coverage from 2004 to 2006. This study found that hospitals with high percentages of Medicaid patients had worse performance in 2004 and had significantly smaller improvements over time than those with low percentages of Medicaid patients.</td>
</tr>
<tr>
<td></td>
<td>Ryan (2013)</td>
<td>Examined incentive payments in the first year of Medicare’s Hospital-Value Based Purchasing (HVBP) program. Using a Disproportionate Share Hospital (DSH) index, Ryan found that hospitals with a higher DSH index value had significantly lower Medicare payment adjustments (p &lt; 0.01) in the first year of HVBP, which resulted in a significantly more negative expected financial impact.</td>
</tr>
<tr>
<td></td>
<td>Friedberg et al. (2010)</td>
<td>In the primary care setting, P4P payments were simulated to 438 Massachusetts practices serving higher and lower shares of patients from vulnerable communities. Various preventive care categories were examined, ranging from cancer to diabetes screening, for practices with low to high racial/ethnic vulnerability and economic vulnerability. The authors found no instance in which practices serving the most vulnerable patient catchment areas would receive higher payments than other practices.</td>
</tr>
<tr>
<td><strong>Pathway 2: “Color-blind” quality improvement programs</strong></td>
<td>Gornick et al. (1996)</td>
<td>Linked 1990 census data on median income according to ZIP Code with 1993 Medicare administrative data for 26.3 million beneficiaries 65 years of age or older (24.2 million whites and 2.1 million blacks) Calculated age- and sex-adjusted rates of various diagnoses and procedures according to race and income and computed black to white ratios. The study concluded that race had substantial effects on use of services among Medicare beneficiaries and that Medicare coverage alone is not sufficient to promote effective patterns of preventive health service use by all beneficiaries.</td>
</tr>
</tbody>
</table>
|                                                                              | Mead et al. (2008)        | Used a variety of sources ranging from national surveys to single-site studies to demonstrate differences in experiences of services (e.g. black women were more likely to receive inadequate communication of their }
<table>
<thead>
<tr>
<th>Pathway 3: “Teaching to the test” might disproportionately affect minorities</th>
<th>Casalino, Alexander, et al. (2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conducted a national survey of 556 general internists’ views on pay-for-performance and public reporting of quality scores. A majority believed that “teaching to the test” would occur.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Doran et al. (2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Through conducting a longitudinal analysis of achievement rates for 42 activities (23 included in the incentive scheme and 19 not included), investigated whether the UK’s incentive scheme led general practitioners to neglect activities not included in the scheme. Found that aspects of quality that were incentivized did better than predicted from previous trends, whereas those not incentivized fell below their trend lines.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Campbell et al. (2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied the UK’s QOF. Found that because patient experience was only 2 of 146 incentivized metrics, patient valued aspects of quality did not have success.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Eggleton et al. (2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed a taxonomy of measures used in the Merit-Based Payment System (MIPS) to assess what measurement gaps exist if a more specific focus on health disparities is to occur. By applying a theoretical model of health care quality as a coding matrix, the authors used a directed content analysis approach to categorize individual MIPS measures. Found that most MIPS measures related to aspects of clinical effectiveness,</td>
<td></td>
</tr>
</tbody>
</table>

mammography screening results compared with white women, especially if the results were abnormal).

A longitudinal study of 58,700 hemodialysis patients from 1993 through 2000 used a one-size-fits-all approach for improving hemodialysis dosing for patients with end-stage renal disease, without the recognition that disease burden and treatment choices may differ across race and ethnicity. Led to an overall 40 percent improvement in hemodialysis dosing across the 8-year study period, but did not change the disparity between black and white patients in anemia management and nutritional status.

A multisite randomized clinical trial of 1,801 older adults comparing collaborative care for depression with treatment as usual in primary care had an inclusive approach by recognizing that disease burden and treatment choices were different across race and ethnicity improved depression care. The use of antidepressants by minority and nonminority elderly patients improved and the gap between minorities and nonminorities closed, even though minorities had more severe disease and less antidepressant use at the beginning of the intervention.
whereas few, if any, related to aspects of patient experience or interpersonal care.

**Pathway 4: Avoiding patients perceived as likely to lower quality scores**

McDonald & Roland (2009) 40 primary care physicians (20 from California and 20 from the UK) were interviewed about the unintended consequences of P4P. A common theme that emerged included “Threats to the Ongoing Patient-Physician Relationship.” Californian physicians affiliated with the largest financial incentives expressed frustration about patients who refused to comply with their advice and the inability to exclude individual patients from performance calculations. As a result, some physicians reported behaviors such as forced disenrollment of noncompliant patients. On the other hand, English physicians, who were allowed to exclude or file an “exception report,” did not report feeling resentment or resorting to the methods used by the Californian physicians.

Ryan (2010) Using 2000 to 2006 Medicare data, Ryan tested for differences in the conditional probability of receiving care at Premier Inc. Hospital Quality Incentive Demonstration (PHQID) hospitals, which utilized pay-for-performance (P4P) programs, for acute myocardial infarction (AMI), heart failure, and pneumonia before and after implementation of the PHQID between white and minority patients. The author also tested for differences in the conditional probability that white and minority patients diagnosed with AMI received coronary artery bypass graft (CABG) in hospitals participating and not participating in the PHQID before and after the implementation of the PHQID. Found little evidence that the PHQID reduced access for minority patients. Only “Other Race” beneficiaries had a significant reduction (p < .05) in adjusted admissions to PHQID hospitals in the postperiod, and this was only for AMI. Only marginally significant (p < .10) evidence of a reduction in CABG was found, and this also only occurred for Other Race beneficiaries.

Petersen et al. (2017) Evaluated the effect of a 2007 to 2009 P4P intervention on the quality of hypertension care provided by Veteran Affairs physicians and their primary care panels to black patients and whether it produced risk selection. Using a nested study within a cluster randomized controlled trial of three types of financial incentives and no incentives (control), risk selection was measured by comparing the proportion of patients who switched providers, patient visit frequency, and panel turnover. The authors found that this P4P intervention did not produce risk selection,
as there was no difference between intervention and controls in the proportion of patients who switched providers, visit frequency, or panel turnover.

Dalton et al. (2011) Three cross-sectional analyses were conducted using data from the electronic medical records of all patients with diabetes registered in 23 general practices in Brent, North West London between 2004 to 2005 and 2006 to 2007. Patterns of exclusions were examined for three intermediate outcome indicators (HbA1c, blood pressure, and cholesterol). Study results indicated that black and South Asian patients were more likely to be excluded from the HbA1c indicator than White patients [adjusted odds ratio = 1.64 (1.17–2.29) in 2005 to 2006]. Furthermore, patients registered with practices in deprived areas were consistently exception reported at higher levels.

Table 2: Future Research Agenda Examples

<table>
<thead>
<tr>
<th>Causal Pathways</th>
<th>Potential Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pathway 1:</strong> Reduction in income for physicians in poor minority communities</td>
<td>This pathway is fairly well studied, and we know that the practices most impacted by this pathway are those that are smaller, have limited resources, and have sicker patients. However, it would be helpful to see how different P4P intensities impact this pathway. Would high-intensity P4P programs, such as those in the UK, be impacted significantly more than low-intensity P4P programs, such as those in the U.S? Furthermore, given current policy, another potential study could compare practices before and after the aforementioned 2020 MIPS payment adjustment policy.</td>
</tr>
<tr>
<td><strong>Pathway 2:</strong> “Color-blind” quality improvement programs</td>
<td>This pathway has a lack of evidence in the P4P sphere, but the outside P4P literature tells us that “color-blind” quality improvement programs are occurring. Would “color-blind” programs have a greater impact when racial/ethnic minority patients also have comorbidities, as these patients would require even greater appointment planning and time during appointments? In answering this question, a potential study could examine whether “color-blind” P4P approaches occur more often for patients with no or few comorbidities versus many.</td>
</tr>
<tr>
<td><strong>Pathway 3:</strong> “Teaching to the test” might disproportionately affect minorities</td>
<td>“Teaching to the test” is occurring, and we know current metrics do not reflect those found to be most relevant for racial/ethnic minority patients. Potential metrics that would be relevant should have their effectiveness tested through studies. However, due to smaller and larger clinics having different capacities (e.g. infrastructure and patient sample size), metrics studied should be carefully selected for each practice context. For instance, in a smaller clinic setting, such as a practice with three to five physicians, a measure could be created to see if physicians ask all their patients (to ensure a large enough sample size) about conditions or social circumstances that</td>
</tr>
</tbody>
</table>

36
have a greater impact on minority households, such as stress from perceived discrimination. In contrast, a HIV measure could be created and measured in a larger practice setting, such as a hospital, with greater tracking and technological resources.

Pathway 4: Avoiding patients perceived as likely to lower quality scores

This pathway needs more conclusive evidence. An example of a future study that could help add to the literature includes assessing avoidance of white hypertensive patients versus minority hypertensive patients in small group practices and in large group practices. (Hypertension should be an appropriate condition for both types of practices, as it is highly prevalent among all racial/ethnic groups. If some small practices do not have a large enough sample size of minority patients, results among clinics could perhaps be aggregated.) Petersen et al. already conducted a study for risk selection in black hypertensive patients. However, it was in a larger Veterans Affair hospital-based setting. Therefore, testing this pathway in a smaller setting could be helpful. Furthermore, it could be studied whether there are differences in avoidance for individuals with just hypertension versus individuals with hypertension and other comorbidities.

Figure 1 (Gray et al., 2018)
Figure 2 (Gray et al., 2018)

Figure 20. Number and percentage of quality measures with disparity at baseline for which disparities related to race and ethnicity were improving, not changing, or worsening (2000 through 2014-2016)

Key: n = number of measures; AI/AN = American Indian or Alaska Native; NHPI = Native Hawaiian/Pacific Islander.

Figure 3: Race/ethnicities and percentage with coronary heart disease (CHD) (Braveman et al., 2010)