Disparities and P4P

- Disparities in health and health care continue to exist
- Pay-for-Performance (or P4P)
  - is the practice of rewarding providers to meet quality goals and to improve outcomes of care, rather than paying for the volume of services they provide.
  - has become increasingly common among both private and public payers
- With one exception*, P4P programs (and their close cousin, public reporting) have not been designed to measure and reward reductions in disparities
- Experts have raised a number of concerns about P4P and disparities

* P4P for disparities reduction in the Massachusetts MassHealth program, 2007-2010
The Princess Bride and the Fire Swamp

**Buttercup:** We'll never succeed. We may as well die here.

**Westley:** No, no. We have already succeeded. I mean, what are the three terrors of the Fire Swamp? One, the flame spurt - no problem. There's a popping sound preceding each; we can avoid that. Two, the lightning sand, which you were clever enough to discover what that looks like, so in the future we can avoid that too.

**Buttercup:** Westley, what about the R.O.U.S.’s?

**Westley:** Rodents Of Unusual Size? I don’t think they exist.

P4P and Disparities: Seven Terrors of The Fire Swamp

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. A national P4P program wouldn’t have any impact
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?
No Data means
No Information means
No Improvement

- Race, ethnicity, and language data are needed to:
  - Stratify quality performance metrics
  - Organize and focus quality improvement and disparity reduction initiatives
  - Track progress over time, locally and as a nation

Romana Hasnain-Wynia, PhD

“How concerned would you be that this data could be used to discriminate against patients?”

<table>
<thead>
<tr>
<th>Concern Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not concerned at all</td>
<td>34%</td>
</tr>
<tr>
<td>A little concerned</td>
<td>15%</td>
</tr>
<tr>
<td>Somewhat concerned</td>
<td>20%</td>
</tr>
<tr>
<td>Very concerned</td>
<td>31%</td>
</tr>
</tbody>
</table>

“It is important for hospitals and clinics to conduct studies to make sure that all patients get the same high-quality care regardless of their race or ethnic background”

<table>
<thead>
<tr>
<th>Opinon</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>93%</td>
</tr>
<tr>
<td>Somewhat agree</td>
<td>4%</td>
</tr>
<tr>
<td>Unsure</td>
<td>2%</td>
</tr>
<tr>
<td>Somewhat disagree</td>
<td>1%</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>0%</td>
</tr>
</tbody>
</table>


Tell People Why You are Asking

“Now I would like you to tell me your Race and Ethnic Background. We use this to review the treatment patients receive and make sure everyone gets the highest quality of care.”

- Of the participants who were not completely comfortable reporting their race and ethnicity
  - 25.0% said that the quality statement made them somewhat more comfortable
  - 25.6% said the quality statement made them much more comfortable

The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. A national P4P program wouldn’t have any impact
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?

The Rising Tide Lifts All Boats Question:
Does quality improvement necessarily lead to reductions in disparities?

Scenario 1: Improvement in both groups; gap narrows
The Rising Tide Lifts All Boats Question:
Does quality improvement necessarily lead to reductions in disparities?

Scenario 2: More improvement in whites than blacks; gap widens

The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. A national P4P program wouldn’t have any impact
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?
Did Racial-Ethnic Disparities Get **Better or Worse** Between 2000-2010?

**Answer: Both!**

- **The B/W difference** got **better** over time (from 15➔10)
- **The B/W ratio** got **worse** over time (from 1.6➔2.0);
Recommendation – this one’s easy

- Absolute and relative changes in disparities can yield different conclusions on whether or not gaps are closing
- Both types of statistics should be calculated, and if they lead to conflicting conclusions, both should be presented, allowing readers to make their own interpretation

Did Racial-Ethnic Disparities Get **Better or Worse** Between 2000-2010?
The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. **P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities**
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. A national P4P program wouldn’t have any impact
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?

Ongoing Concerns

- Traditional P4P programs do not appropriately account for differences in the medical or social complexity of patients, and as a result may widen disparities in healthcare because they may allow the “rich” to get richer while the “poor” get poorer
- Ability to reduce disparities depends on the SES of the patients

Thanks in part to Alyna T. Chien, MD, MS
IHA’s P4P Program: Physician Organizations Located in Lower Socioeconomic Status Areas Scored Lower on Pay for Performance Measures

The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. **P4P creates incentives in the wrong place** *(Looking for your keys under the lamp post)*
6. A national P4P program wouldn’t have any impact
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?
Are racial-ethnic disparities primarily a matter of unequal treatment by providers?

The “Usual” Explanations of Disparities

- **It’s the doctor’s fault**
  - Beliefs/stereotypes about behavior or health of patients
  - Bias/Prejudice

- **It’s the patient’s fault**
  - Patient preferences, including treatment refusal, clinical presentation of symptoms

So what might be behind the divergent views???
What the Doctors Say

I treat all of my patients the same...

CHALLENGE

Who You Are?: Patient-Centered Care

Where You Go?: Quality differences in various settings

Thanks to Romana Hasnain-Wynia
Evidence of a “between” phenomenon: Bottom performing hospitals had a much higher percentages of minority patients compared with top performing hospitals.

![Bar chart showing minority patients percentages between top and bottom performers.]

Data from 123 University HealthSystem Consortium hospitals, Q3 2002 to Q1 2005, n=320,970 adult patients.


Evidence of both: Disparities in Hospital Quality Alliance (HQA) Measures is a Mix of Between and Within Phenomena.

<table>
<thead>
<tr>
<th>% White - % Minority Who Obtain Recommended Care</th>
<th>Unadjusted</th>
<th>Ind'l Char'ics Adjustment*</th>
<th>Ind'l and Hospital Char'ics**</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF-Discharge Instructions</td>
<td>9.6</td>
<td>9.1</td>
<td>0.5*</td>
</tr>
<tr>
<td>HF-Smoking Cessation</td>
<td>8.4</td>
<td>8.2</td>
<td>3.5@</td>
</tr>
<tr>
<td>AMI-Smoking Cessation</td>
<td>9.0</td>
<td>9.1</td>
<td>3.8@</td>
</tr>
<tr>
<td>CAP- Antibiotic therapy</td>
<td>9.8</td>
<td>8.3</td>
<td>6.4@</td>
</tr>
</tbody>
</table>

* - age, sex, payer, severity of illness
** Includes a dummy variable for each hospital
@ - p<.05

Data from 123 University HealthSystem Consortium hospitals, Q3 2002 to Q1 2005, n=320,970 adult patients.

Hasnain-Wynia, R., Baker, DW, Nerenz, DR, Weissman JS. "Disparities in Health Care Are Driven by Where Minority Patients Seek Care: An Examination of the Hospital Quality Alliance Measures: Archives of Internal Medicine June 25, 2007 627:1233-1239."
Different policy implications

- Patient-Centered ➜ Cultural competency training
- System ➜ Health policy (payment), housing, transportation

The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. **A national P4P program wouldn’t have any impact**
7. Hasn’t this been tried anywhere? In Massachusetts! What happened?
Devil in the Details: The Design of Pay-For-Performance Programs for Reducing Disparities: What do the Data Tell Us?


P4P (and public reporting) Design Issues

- Minimum case requirements
  - May eliminate many or most hospitals from eligibility for disparities-based measurement and therefore participation in P4P
- Ranking methods matter and are still in development
  - “Quality Improvement” vs “Disparities reduction”
- OUR STUDY
  - Part I: Identify the number of U.S. hospitals that meet the 30/30 threshold
  - Part II: Simulate results of a national P4P disparities reduction program using different ranking methods and examine the potential impact on nationwide disparities
Data

- CMS National Hospital Quality Alliance
  - AMI, HF, PN
  - All payer
  - CY 2005
- 2.3 million discharges from 4,450 non-federal hospitals.


Most Hospitals Do Not have the Minimum # of Cases for Disparity P4P

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Hospitals</th>
<th>Hospitals w/ Any Cases</th>
<th>Hospitals w/ 30/30 Cases*</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td>3999</td>
<td>1443</td>
<td>36%</td>
<td></td>
</tr>
<tr>
<td>CHF</td>
<td>4355</td>
<td>1946</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>PNE</td>
<td>4425</td>
<td>2353</td>
<td>53%</td>
<td></td>
</tr>
</tbody>
</table>

* -- at least 30 white and 30 minority cases
But... Hospitals that Met the 30/30 Threshold Treated Most of the Minority Patients

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Minority Patients Treated...</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In Hospitals w/ Any Cases</td>
<td>In Hospitals w/ 30/30 Cases</td>
</tr>
<tr>
<td>AMI</td>
<td>92,000</td>
<td>84,000</td>
</tr>
<tr>
<td>CHF</td>
<td>241,000</td>
<td>223,000</td>
</tr>
<tr>
<td>PNE</td>
<td>159,000</td>
<td>152,000</td>
</tr>
</tbody>
</table>

Simulation Methods

- Calculate scores for each race-ethnicity within each hospital and nationally
  - Quality Scores
  - Disparity Scores
- Rank the hospitals
- Simulate “success”
  - Make the bottom half look like the top
  - Re-calculate national scores
- Assess two national impacts
  - Increase in overall quality scores
  - Reduction in disparities scores
Simulated Changes in National Quality and Disparity Scores for Pneumonia Using Two Methods to Rank Hospitals – Plus a “Combined” Method

Change in Score After Simulation

Weissman JS, et al 2012

The Fire Swamp of P4P and Disparities

1. We don’t collect race-ethnicity data and patients don’t think we should
2. The rising tide argument doesn’t work - “Color-blind” quality improvement programs may help non-minorities more
3. Changes in disparities depend on how you measure them
4. P4P creates an uneven playing field: Possible reduction in income for providers in poor minority communities
5. P4P creates incentives in the wrong place (Looking for your keys under the lamp post)
6. A national P4P program wouldn’t have any impact
7. 

*Hasn’t this been tried anywhere? In Massachusetts! What happened?*
Chapter 58-The Massachusetts Universal Health Care Law Also Mandated Hospital P4P

- Boston Task Force to Eliminate Racial and Ethnic Disparities (2004-5)
- Section 25 of Chapter 58 (2006)
  - MassHealth hospital rate increases contingent on quality standards, including the reduction of racial and ethnic disparities.

MassHealth Hospital Pay for Performance, MassHealth Symposium, 2007
Center for Surgery and Public Health

P4P Measures

1) Clinical Measures - Reward Hospitals to report data by Race/Ethnicity & to reduce differences in clinical processes
2) Structural Measures - Reward Hospitals to improve organizational factors that may reduce racial/ethnic health disparities.
Mixed Reactions from Hospital Community

- Strong stated support for the program’s goal
- Participation required extraordinary effort
- Frustration with the effort required to adapt to the clinical reporting system
- Some measures felt to be “ambiguous”
- Perceived focus on documentation at the expense of quality improvement

Care of minorities is relatively concentrated compared with whites

4 most voluminous hospitals see 20.6% of whites and 33.2% of minorities

Blustein J, Weissman JS, Ryan AM, Doran T, Hasnain-Wynia R. Analysis raises questions on whether pay-for-performance in Medicaid can efficiently reduce racial and ethnic disparities Health Affairs 2011; 30(6):1165-1175
Differences among hospitals were small and could not be reliably estimated due to small numbers of minorities

Blustein J, Weissman JS, Ryan AM, Doran T, Hasnain-Wynia R. Analysis raises questions on whether pay-for-performance in Medicaid can efficiently reduce racial and ethnic disparities Health Affairs 2011; 30(6):1165-1175
So, Can P4P disparities program have a happy ending?

Or Death by R.O.U.S.’s?
Maybe....
In some cases....
If carefully designed....

Possible paths through the fire swamp - 1

- Data collection
  - It can be done if presented in the context of quality
  - Quality dashboards should be stratified by race-ethnicity

- Rising Tides vs Targeted Improvement
  - Instead of choosing one vs the other ➔ Focus on underlying factors, e.g., if it’s a language problem, get interpreters; if it’s a poverty problem, help patients who are poor; if it’s an access problem, then focus on getting access to services for everyone and provide a navigator to lead the way
Possible paths through the fire swamp - 2

- Disparities measurement
  - Check that different methods lead to similar conclusions
  - Understand the underlying numbers
  - Be transparent
- Disparities dashboards on the other hand make a lot of sense

Possible paths through the fire swamp - 3

What about P4P to reduce disparities?

- First, look at the data (make sure there are big enough disparities to address for the conditions of interest using available measures)
- Traditional P4P programs should reward improvement as well as threshold attainment
- Sample size problems should be addressed up front
- Not all statistically significant differences are meaningful
- The “between” problem should be examined along with the “within” problem.
Possible paths through the fire swamp - 4

Risk adjustment

- Should *traditional* P4P programs risk adjust for race-ethnicity? Or should they stratify?
- Should *disparities* P4P programs risk adjust for SES?
- What about something completely different?
  - Risk adjustment payments not outcomes

In Sum - no simple story

Despite growing interest in using P4P to reduce disparities in healthcare, we may not be quite ready yet to implement the idea in a high stakes program.

Not only do we need to know more about measures that are “disparities-sensitive”, but how to select measures that are ready to have an impact on clinical practice, and how to represent differences in a statistically meaningful and policy-relevant way.
David Satcher, M.D., Ph.D.
Former Surgeon General of the U.S.

“In order to eliminate disparities in health, we need leaders who care enough, know enough, will do enough and are persistent enough.”

British Reporter Interviewing Mohandas Gandhi During the Indian Revolution:

- Reporter: “Mr. Gandhi, what do you think of Western civilization?”

- Gandhi (after a pause): “I think it would be a very good idea.”

Thanks to Larry Casalino for quote
The IOM Definition: Differences, Disparities & Discrimination

Clinical Need & Appropriateness, Patient Preferences

Healthcare Systems & Legal / Regulatory Systems

Disparities: Bias, Stereotyping, & Uncertainty

IOM, 2002