

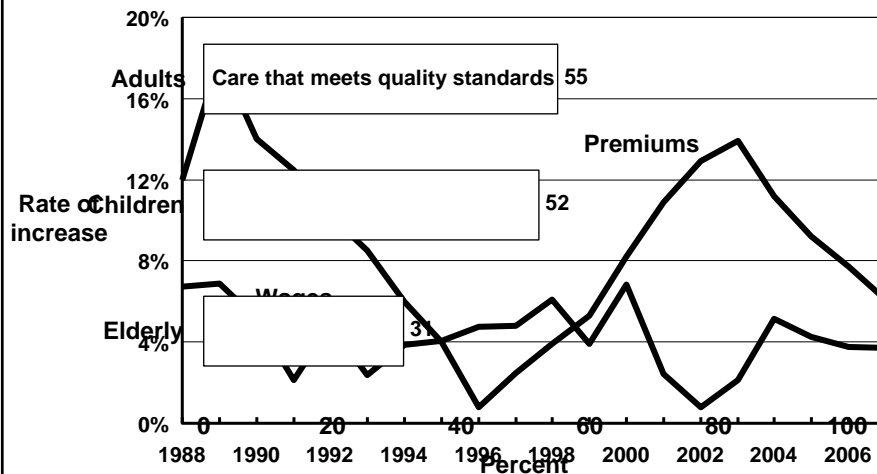


HEALTH

# Issues in Constructing Efficiency (Cost of Care) Performance Metrics on Physicians

Elizabeth A. McGlynn, Ph.D.  
Associate Director, RAND Health  
September 16, 2008

## Health Insurance Premiums Are Increasing Faster than Wages



RAND

McGlynn WHQC 2 09/2008

## ***Many Policy Interventions Have Been Tried***

Capitation  
Managed Care  
Disease Management  
Guidelines  
Consumer-Directed Health Care  
Gatekeepers  
Transparency  
Pay-for-Performance

RAND

McGlynn WHQC 3 09/2008




## ***Involving Consumers in Health Care Decisions Could Move the Market***

- **Patients are the engine of change**
  - Freedom to choose providers
  - Information to make better choices
  - Pay more for expensive care
- **These “value-based” decisions will motivate providers to change**

RAND

McGlynn WHQC 4 09/2008

## ***A New Role for Patients***

	<b>Information</b>		<b>Financial Incentive</b>
	<b>Quality</b>	<b>Costs</b>	<b>Tiering</b>
	☆☆☆	\$	<u>High value</u> \$10 Co-payment
	☆☆☆	\$\$	<u>Average value</u> \$20 Co-payment
	☆	\$\$\$	<u>Low value</u> \$30 Co-payment

RAND

McGlynn WHQC 5 09/2008

## ***RAND Project Goals***

- Open the “black box” and assess the methodology of profiling
- Identify which methodological choices matter
- Highlight what that means for policy

RAND

McGlynn WHQC 6 09/2008

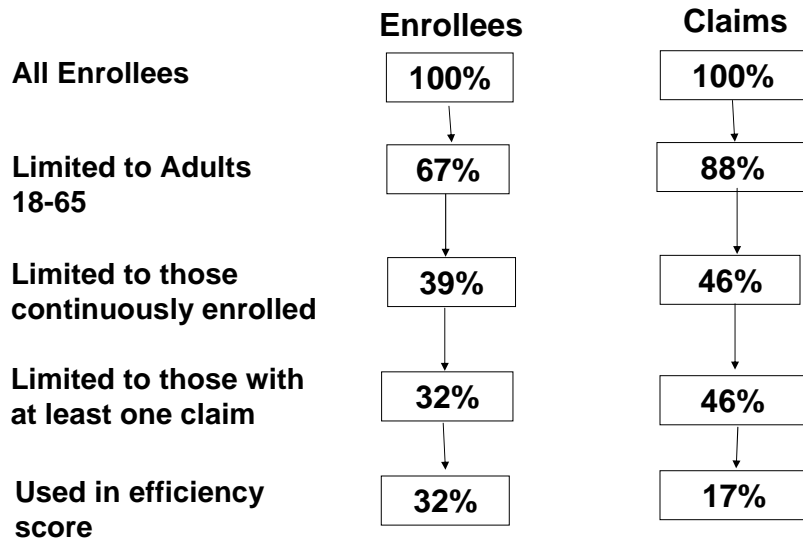
## ***Analytic Data Set***

- **Aggregated claims from 4 Massachusetts health plans**
  - Data from 2003-2005
  - 2.9 million commercial enrollees
  - Adults <65 who were continuously enrolled for two years

RAI 10

McGlynn WHQC 7 09/2008

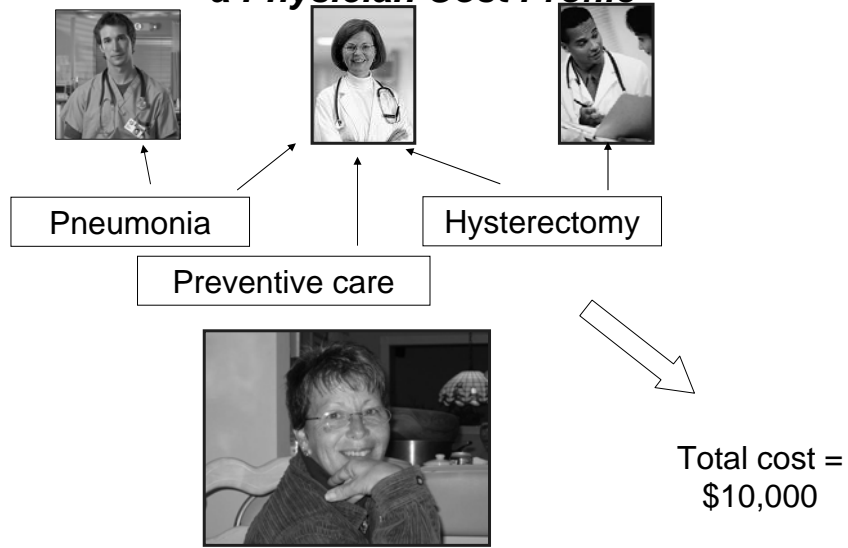
## ***Our Scores Are Based on a Minority of Patients and Claims***



RAI 10

McGlynn WHQC 8 09/2008

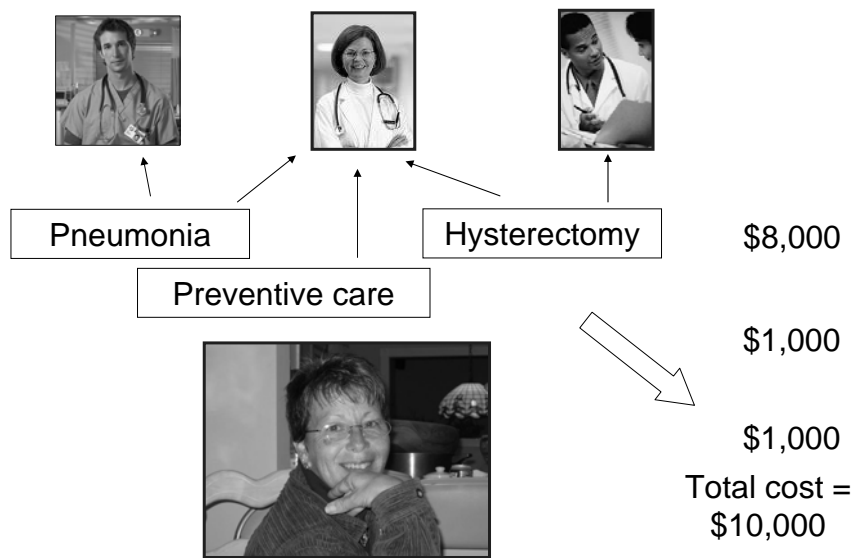
### Common Approach to Creating a Physician Cost Profile



RAND

McGlynn WHQC 9 09/2008

### Creating a Physician Cost Profile



RAND

McGlynn WHQC 10 09/2008

## Comparing Costs for Each Episode



	Care assigned to Dr. Matthews	Average for that condition
Hysterectomy	\$8,000	\$14,000
Preventive care	\$1,000	\$500
Pneumonia	\$1,000	\$600
<b>TOTAL</b>	<b>\$10,000</b>	<b>\$15,100</b>

RAND

McGlynn WHQC 11 09/2008

## We've Examined Several Methodological Choices Required to Create Physician Profiles

Aggregation	Does combining data across health plans increase opportunities to profile?
Attribution	Which physician is assigned responsibility?
Reliability	Is there a minimum reliability for a physician profile?
Classification	How are physicians assigned to categories of performance?
Metric	How is the metric constructed?
Level of Analysis	Should the level of analysis be individual physicians, practices, or groups?
Standardized Costs	Should actual reimbursements to the provider or standardized costs be used?

RAND

McGlynn WHQC 12 09/2008

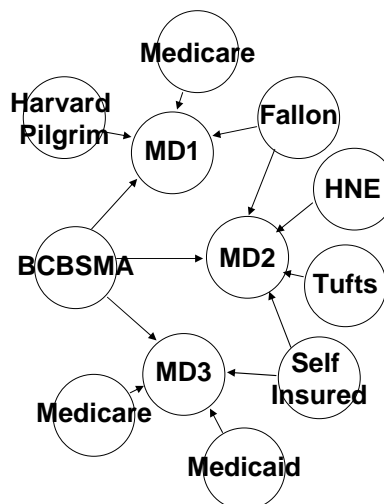
## ***We've Examined Several Methodological Choices Required to Create Physician Profiles***

<b>Aggregation</b>	Does combining data across health plans increase opportunities to profile?
<b>Attribution</b>	Which physician is assigned responsibility?
<b>Reliability</b>	Is there a minimum reliability for a physician profile?
<b>Classification</b>	How are physicians assigned to categories of performance?
<b>Metric</b>	How is the metric constructed?
<b>Level of Analysis</b>	Should the level of analysis be individual physicians, practices, or groups?
<b>Standardized Costs</b>	Should actual reimbursements to the provider or standardized costs be used?

RAND

McGlynn WHQC 13 09/2008

## ***Physicians Have Multiple Contracts***



**So, having enough information matters:  
Aggregation**

RAND

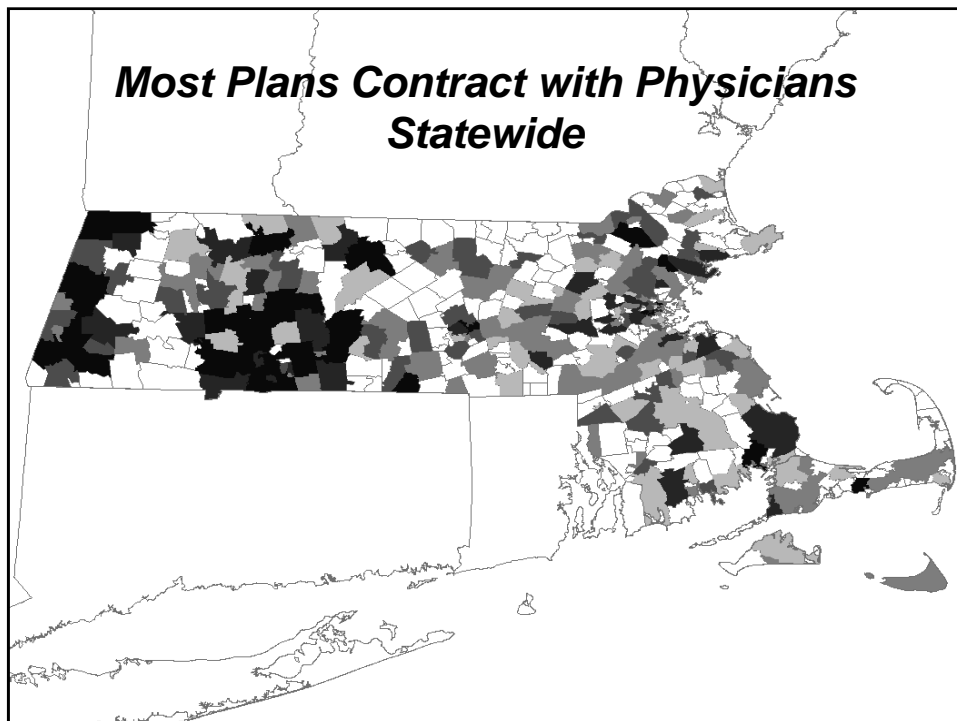
McGlynn WHQC 14 09/2008

## ***Some Challenges in Aggregation***

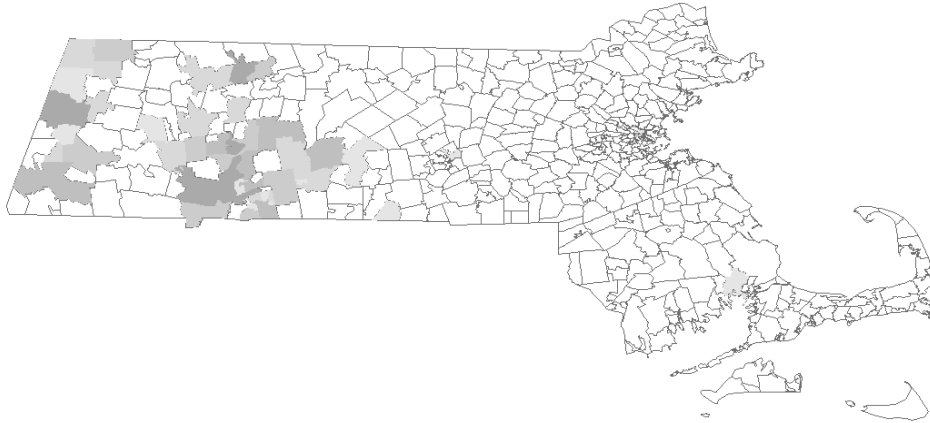
- **Creating a master directory**
- **Figuring out who is who**
- **Deciding on the specialty**
- **Putting codes together**
- **Figuring out how much each piece costs**

RAT 10

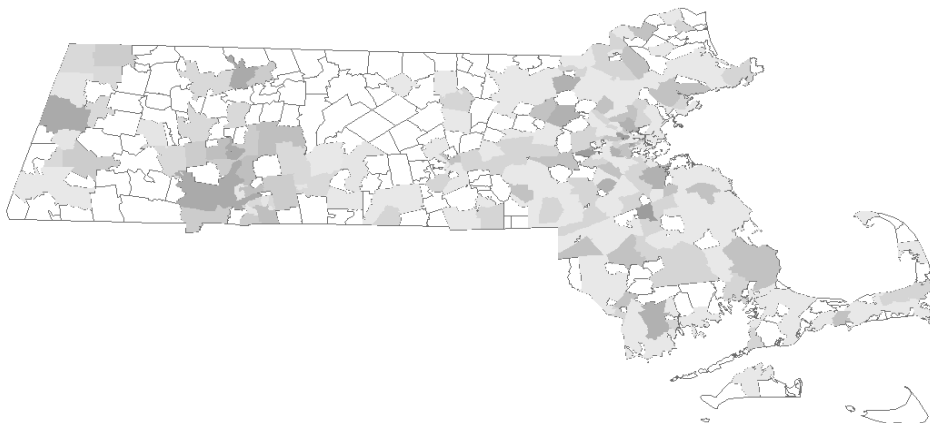
McGlynn WHQC 15 09/2008



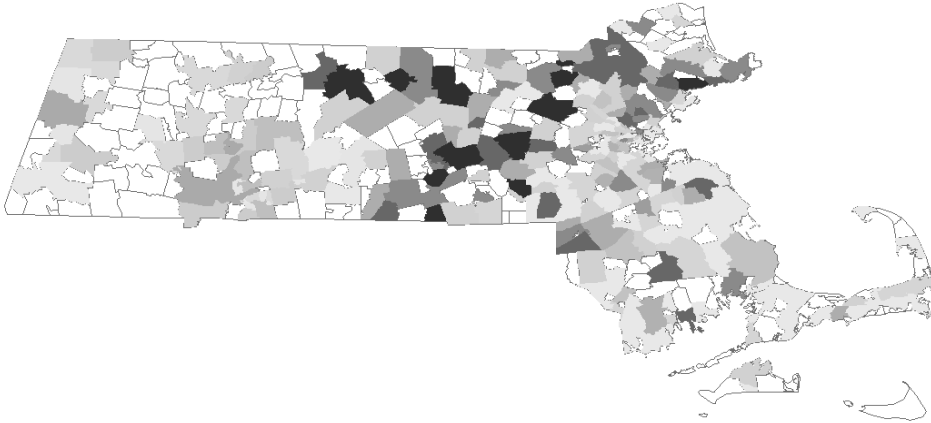
**But Utilization Is Concentrated in  
One Area of the State**



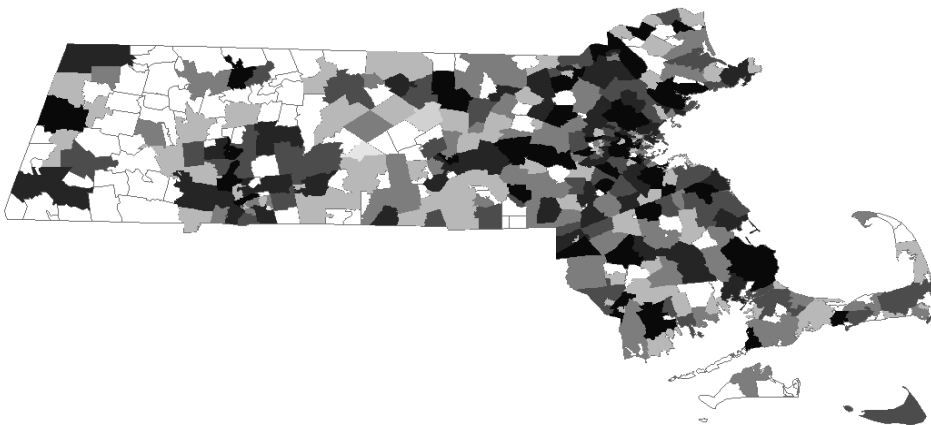
**So Aggregating These Plans May Not  
Increase Sample Sizes**



**Or Any Two of These Plans**



***But Aggregating Data with this Purchaser  
Increases Number of Observations***



## ***We've Examined Several Methodological Choices Required to Create Physician Profiles***

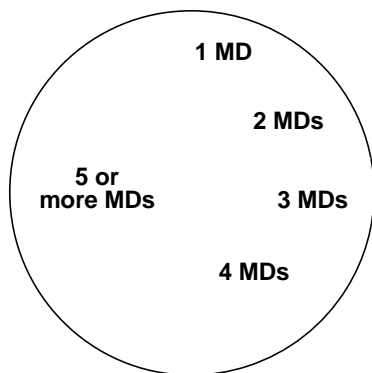
<b>Aggregation</b>	Does combining data across health plans increase opportunities to profile?
<b>Attribution</b>	Which physician is assigned responsibility?
<b>Reliability</b>	Is there a minimum reliability for a physician profile?
<b>Classification</b>	How are physicians assigned to categories of performance?
<b>Metric</b>	How is the metric constructed?
<b>Level of Analysis</b>	Should the level of analysis be individual physicians, practices, or groups?
<b>Standardized Costs</b>	Should actual reimbursements to the provider or standardized costs be used?

RAND

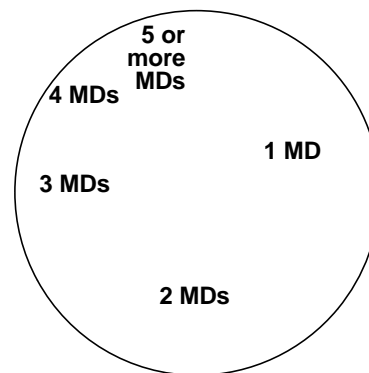
McGlynn WHQC 21 09/2008

## ***Why Is Attribution Important?***

**Patient**



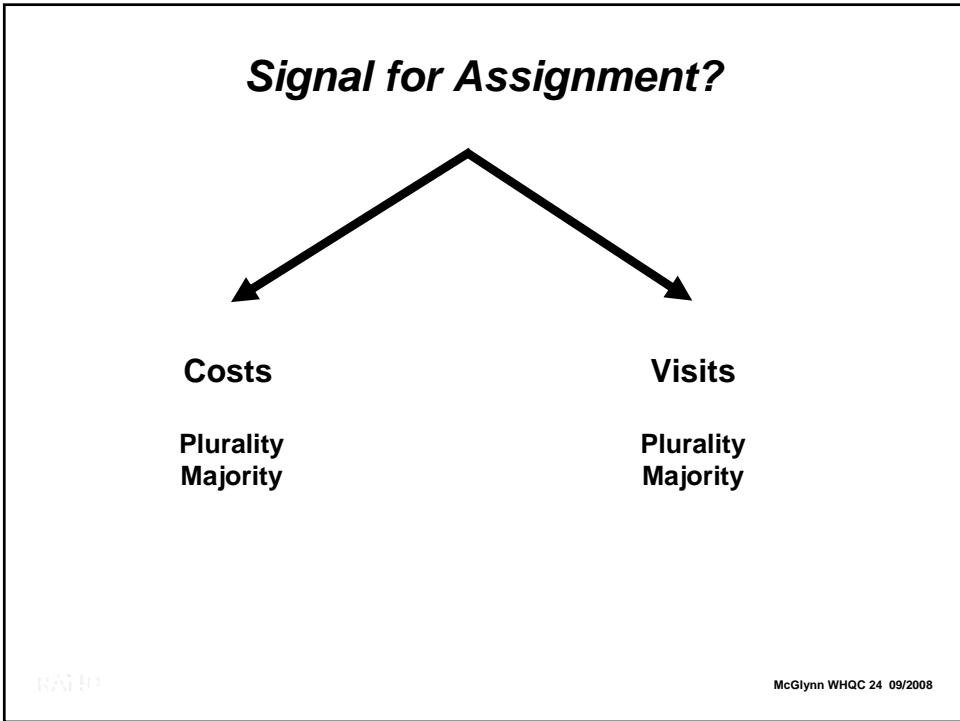
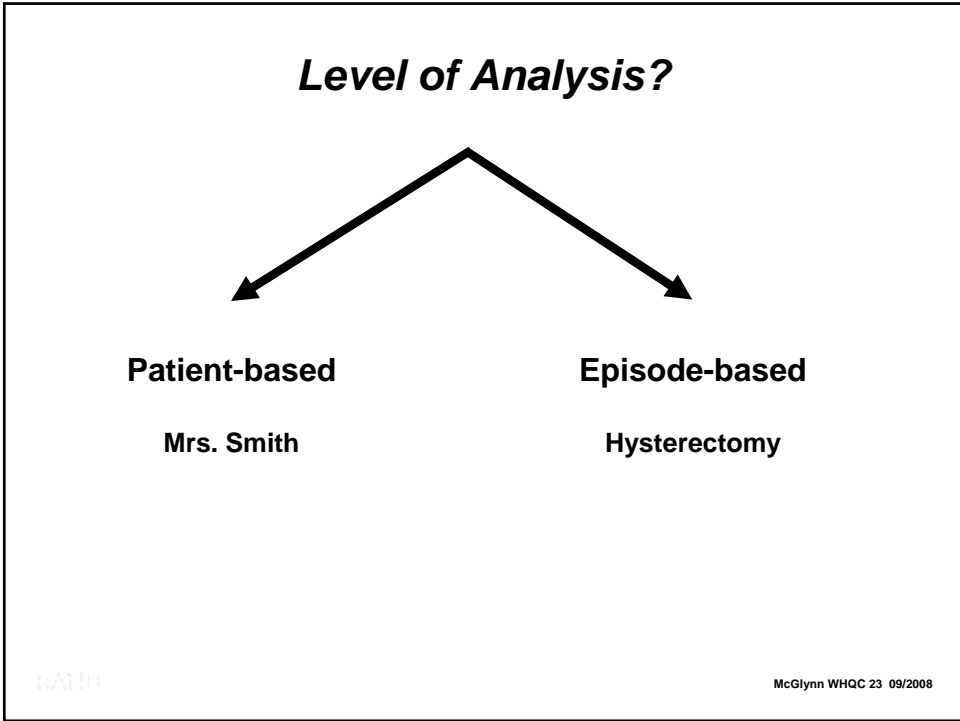
**Episode**



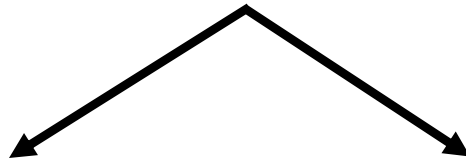
**Which physician is responsible for care?**

RAND

McGlynn WHQC 22 09/2008



## One or Multiple Physicians?



Single Physician

Dr. Matthews

Multiple Physicians

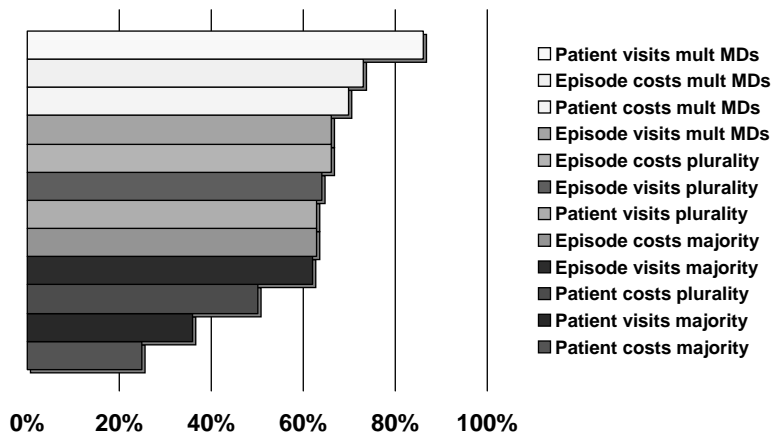
Dr. Matthews &  
Dr. Cutter

We evaluated 12 attribution rules

RAT 10

McGlynn WHQC 25 09/2008

## Proportion of Episodes That Can Be Attributed Varies by Method



RAT 10

McGlynn WHQC 26 09/2008

***Are the Same Episodes Assigned to the Same Physician under Different Rules?***

<b>Episode-Based</b>	<b>Agreement</b>	<b>Patient-Based</b>
<b>Cost-Plurality</b>		<b>Cost-Plurality</b>
<b>Visits-Plurality</b>		<b>Visits-Plurality</b>
<b>Costs-Majority</b>		<b>Costs-Majority</b>
<b>Visits-Majority</b>		<b>Visits-Majority</b>

RAND

McGlynn WHQC 27 09/2008

***Agreement on Episode Assignment is Moderate to High***

<b>Episode-Based</b>	<b>Agreement</b>	<b>Patient-Based</b>
<b>Cost-Plurality</b>	<b>58%</b>	<b>Cost-Plurality</b>
<b>Visits-Plurality</b>	<b>61%</b>	<b>Visits-Plurality</b>
<b>Costs-Majority</b>	<b>71%</b>	<b>Costs-Majority</b>
<b>Visits-Majority</b>	<b>73%</b>	<b>Visits-Majority</b>

RAND

McGlynn WHQC 28 09/2008

**Physicians Assigned to the Same Category of Performance Most of the Time**

<b>Episode-Based</b>	<b>Same Rating?</b>	<b>Patient-Based</b>
Cost-Plurality	<b>88%</b>	Cost-Plurality
Visits-Plurality	<b>89%</b>	Visits-Plurality
Costs-Majority	<b>89%</b>	Costs-Majority
Visits-Majority	<b>90%</b>	Visits-Majority

RAND

MD assigned under both rules

McGlynn WHQC 29 09/2008

**Physicians Assigned to the Same Category of Performance Most of the Time (v 2)**

<b>Episode-Based</b>	<b>Same Rating?</b>	<b>Patient-Based</b>
Cost-Plurality	<b>73%</b>	Cost-Plurality
Visits-Plurality	<b>73%</b>	Visits-Plurality
Costs-Majority	<b>62%</b>	Costs-Majority
Visits-Majority	<b>62%</b>	Visits-Majority

RAND

MD assigned under either rule

McGlynn WHQC 30 09/2008

## ***Policy Implications***

- **Choice of attribution rules may affect categorization**
  - No one “right” approach
  - Even 12% of physicians whose assignment would change can create challenges
- **Potential use of information should shape choice of attribution rules**
  - Who is expected to use the information and for what purpose?
  - Will physicians’ responses to results vary by the rule selected?
    - Refuse to see certain patients?
    - Withhold care?

RAI 10

McGlynn WHQC 31 09/2008

## ***We’ve Examined Several Methodological Choices Required to Create Physician Profiles***

<b>Aggregation</b>	Does combining data across health plans increase opportunities to profile?
<b>Attribution</b>	Which physician is assigned responsibility?
<b>Reliability</b>	Is there a minimum reliability for a physician profile?
<b>Classification</b>	How are physicians assigned to categories of performance?
<b>Metric</b>	How is the metric constructed?
<b>Level of Analysis</b>	Should the level of analysis be individual physicians, practices, or groups?
<b>Standardized Costs</b>	Should actual reimbursements to the provider or standardized costs be used?

RAI 10

McGlynn WHQC 32 09/2008

$$\text{reliability} = \frac{\sigma^2_{\text{provider-to-provider}}}{(\sigma^2_{\text{provider-to-provider}} + \sigma^2_{\text{error}})}$$

## ***What Is Reliability?***

- **Reliability:** measures how much variation in observed scores is explained by real differences in performance
- **Expressed as a percentage ranging from 0-100%**
  - **0%** means all of the variability is due to measurement error
  - **100%** means all of the variability is due to real differences in performance

## ***The Math***

$$\text{reliability} = \frac{\sigma^2_{\text{provider-to-provider}}}{\sigma^2_{\text{provider-to-provider}} + \sigma^2_{\text{error}}}$$

***Reliability is an attribute of a physician's own cost profile, not the measure overall***

RAND

McGlynn WHQC 35 09/2008

$$\text{reliability} = \sigma_{\text{provider-to-provider}}^2 / (\sigma_{\text{provider-to-provider}}^2 + \sigma_{\text{error}}^2)$$

### ***What Increases Reliability?***

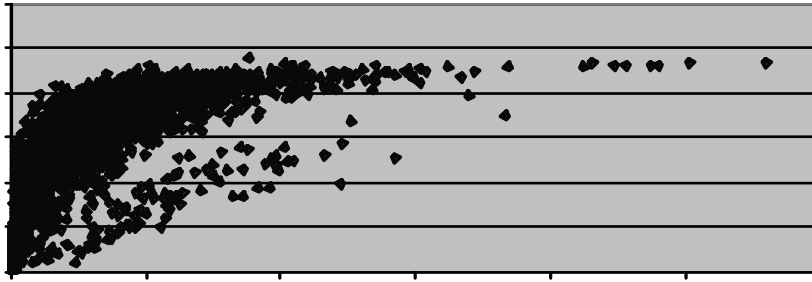
- **A larger number of observations**
  - Eligibility events (quality)
  - Episodes (cost)
- **More real differences in the performance of providers**
- **Less noise in the measures**

**The first component has led groups to develop rules around minimum numbers of observations**

RAND

McGlynn WHQC 36 09/2008

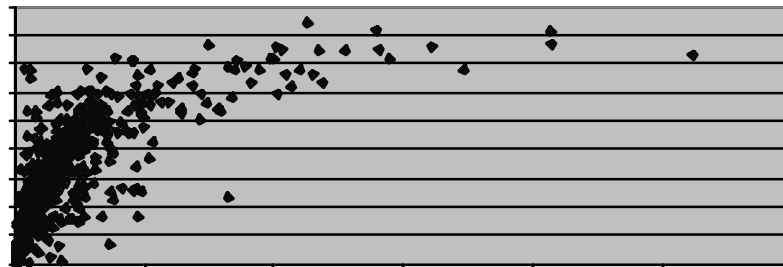
***No Simple Relationship Between Sample Size and Reliability for Cost Data***



RAND

McGlynn WHQC 37 09/2008

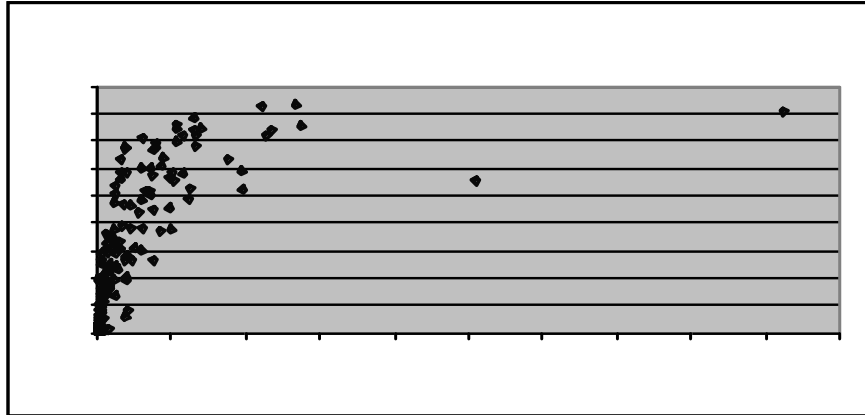
***Similar Pattern for Cardiology***



RAND

McGlynn WHQC 38 09/2008

## *And for Endocrinology*



RAND

McGlynn WHQC 39 09/2008

## *Is there a Minimum Level of Reliability?*

- Psychometricians use a rule of thumb of 90% for drawing conclusions about individuals
- Lower levels (70-80%) are considered acceptable for drawing conclusions about groups
- Choice of level raises questions about the tradeoff between feasibility and scientific soundness

RAND

McGlynn WHQC 40 09/2008

$$\text{reliability} = \frac{\sigma^2_{\text{provider-to-provider}}}{(\sigma^2_{\text{provider-to-provider}} + \sigma^2_{\text{error}})}$$

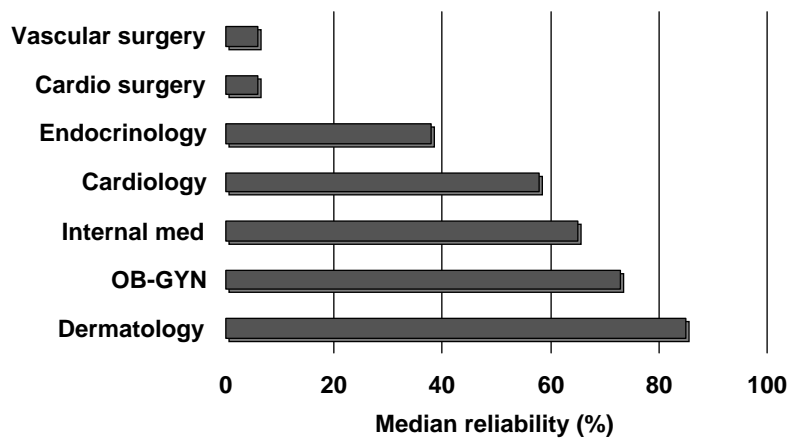
## ***Why Should You Care About Reliability?***

- **Higher reliability increases the likelihood that you will assign a physician to the “right” group**
  - **Using low reliability information to drive behavior change could have undesirable consequences**

RAND

McGlynn WHQC 41 09/2008

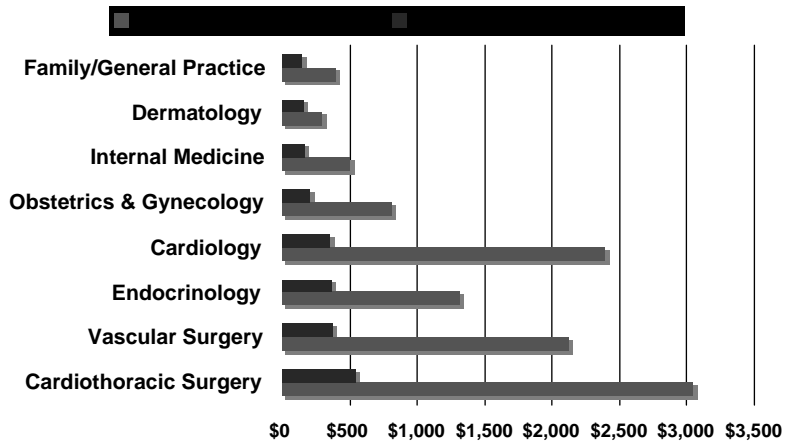
## ***Median Reliability of Physician Cost Profiles Varies Widely by Specialty***



RAND

McGlynn WHQC 42 09/2008

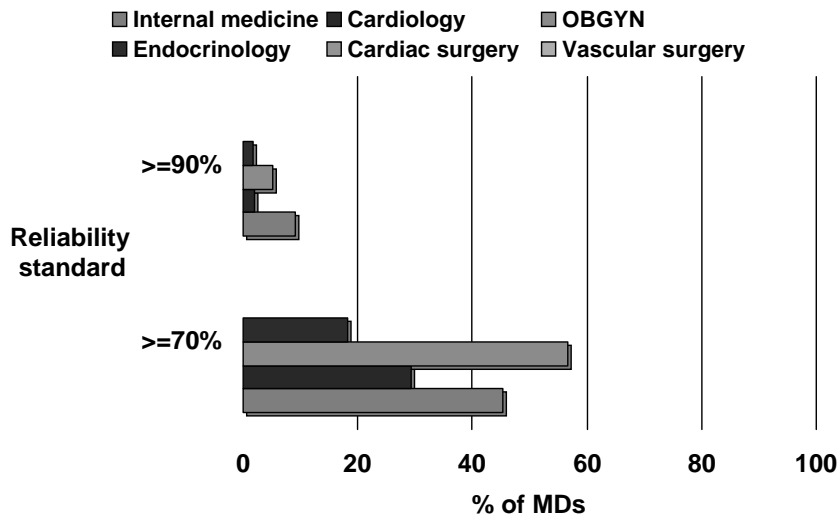
## Variability in Mean and Median Episode Costs by Specialty



RAND

McGlynn WHQC 43 09/2008

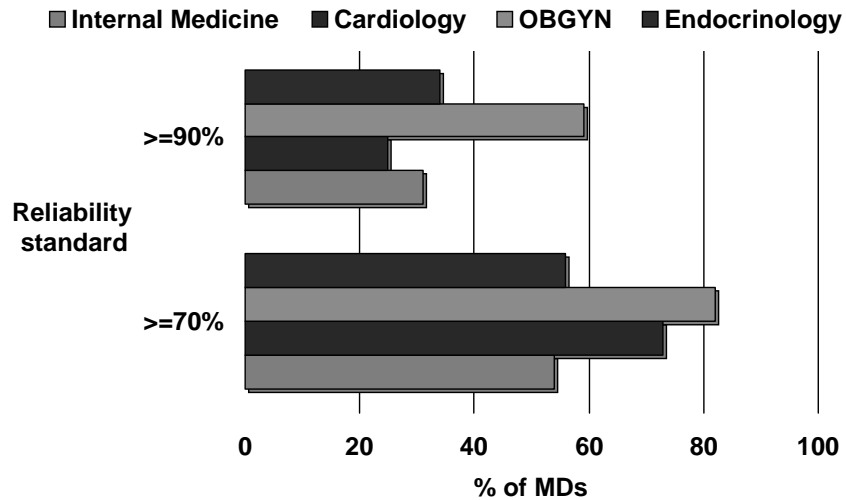
## Proportion of Physicians Whose Cost Profiles Meet Reliability Standards Varies



RAND

McGlynn WHQC 44 09/2008

## Higher Proportion of Physician Quality Profiles Meet Reliability Standards



RAI 10

McGlynn WHQC 45 09/2008

## Policy Implications

- Significant fraction of individual physician profiles are unreliable, leaving the choice between:
  - Not profiling most physicians, or
  - Using unreliable results
- Are low reliability profiles “fair” to physicians?
- Will patients get useful signals from profiles?
- Standards might be different for different purposes
  - Information only
  - Financial incentives
- Caveat: reliability is empirically calculated so these analyses need to be done in each data set

RAI 10

McGlynn WHQC 46 09/2008

## ***We've Examined Several Methodological Choices Required to Create Physician Profiles***

<b>Aggregation</b>	Does combining data across health plans increase opportunities to profile?
<b>Attribution</b>	Which physician is assigned responsibility?
<b>Reliability</b>	Is there a minimum reliability for a physician profile?
<b>Classification</b>	How are physicians assigned to categories of performance?
<b>Metric</b>	How is the metric constructed?
<b>Level of Analysis</b>	Should the level of analysis be individual physicians, practices, or groups?
<b>Standardized Costs</b>	Should actual reimbursements to the provider or standardized costs be used?

RAND

McGlynn WHQC 47 09/2008

## ***Most Applications Use Scores to Categorize Performance***

- **Three levels is a common choice**
  - Above average performance
  - Average performance
  - Below average performance
- **Choice of approach to making assignments to categories**
  - Cutpoints (<25th, 25-75th, >75th percentile)
  - Statistical testing

RAND

McGlynn WHQC 48 09/2008

**Nearly 40% of Physicians Receive Different Assignment Under Two Methods**

		Statistical Testing		
		Low Efficiency	Average Efficiency	High Efficiency
Cut-points	Low Efficiency	7	18	0
	Average Efficiency	0	50	0
	High Efficiency	0	20	5

RAND

McGlynn WHQC 49 09/2008

**Cutpoints Assign More Physicians To Outlier Status than Statistical Testing**

		Statistical Testing			
		Low Efficiency	Average Efficiency	High Efficiency	
Cut-points	Low Efficiency	7	18	0	25
	Average Efficiency	0	50	0	50
	High Efficiency	0	20	5	25
		7	88	5	

RAND

McGlynn WHQC 50 09/2008

## ***Policy Implications***

- **Classification will be different under simple cutpoints versus statistical testing**
- **Cutpoint system will flag some physicians as outliers when there is little statistical evidence that they are different from the average**
- **Given the reliability evidence, this finding should inspire demands for the use of more rigorous tests in classification schemes**
- **Tradeoff between more false positives (cutpoints) and more false negatives (testing)**

RAND

McGlynn WHQC 51 09/2008

## ***Our Work is Continuing...***

<b>Aggregation</b>	<b>Does combining data across health plans increase opportunities to profile?</b>
<b>Attribution</b>	<b>Which physician is assigned responsibility?</b>
<b>Reliability</b>	<b>Is there a minimum reliability for a physician profile?</b>
<b>Classification</b>	<b>How are physicians assigned to categories of performance?</b>
<b>Metric</b>	<b>How is the metric constructed?</b>
<b>Level of Analysis</b>	<b>Should the level of analysis be individual physicians, practices, or groups?</b>
<b>Standardized Costs</b>	<b>Should actual reimbursements to the provider or standardized costs be used?</b>

RAND

McGlynn WHQC 52 09/2008

## ***What Are The High Leverage Decisions?***

- **Aggregation**
  - Context-specific, but increasing number of observations increases likelihood of accurate assessment
- **Attribution**
  - Main difference is between patient and episode based approaches
- **Reliability**
  - This is the most critical problem we've found
- **Classification**
  - Statistical testing superior to cutpoints
- **Validity and actionability remain open challenges**

RAND

McGlynn WHQC 53 09/2008

## ***Where To Go From Here?***

- **Essential to find a way to improve quality and manage costs (as well as ensure access)**
- **There are probably not painless ways to do this**
- **Ideally these decisions will involve all stakeholders**
  - but getting something for nothing probably isn't one of the options
    - So, what is each group willing to give up to ensure value and sustainability in the health system?
    - Transparent, participatory processes are critical for moving forward

RAND

McGlynn WHQC 54 09/2008