

## Mgmt 444

### Quality

Thus far, we have talked about quality largely in the abstract

- We discussed research that measured quality in terms of mortality rates, or service offerings, or even equated quality with cost.

Today we will learn how to think about the many dimensions of quality, how it is measured, and how it is reported

- We will take a look at the ongoing report card movement
- We will try to determine what more needs to be done

Begin with some simple economic theory

Q: What are the potential benefits of report cards?

Before we can discuss quality shopping, we should talk about how to measure it

We can divide the aspects of quality into two broad categories:

- (1) Dimensions that are inherently important (mortality, amenities)
- (2) Those that might affect the first category (equipment, staffing)

It is often difficult to measure dimensions of quality in category (1)

- There are a variety of reasons that we will address later on

We often must rely on dimensions in category (2)

- E.g., the new Medicare Nursing Home Compare lists staffing data

<b><u>Nursing Homes and/or Skilled Nursing Facilities</u></b> <b>for residents needing short or long stay</b>					
<b>Nursing Staff Hours Per Resident Per Day,</b> is the average hours worked by the licensed nurses or nursing assistants divided by total number of residents. The amount of care given to each resident varies.					
	<b>Number of Residents</b>	<b>Licensed RN Hours per Resident per Day</b>	<b>Licensed LPN/LVN Hours per Resident per Day</b>	<b>Total Number of Licensed Nurse Staff Hours per Resident per Day</b>	<b>CNA Hours per Resident per Day</b>
National Average	95.4	30 minutes	42 minutes	1 hour 12 minutes	2 hours 18 minutes
Average in Illinois	104.5	36 minutes	30 minutes	1 hour 6 minutes	2 hours
ABBOTT HOUSE	Not Available - Your State Survey Agency may have the staffing information for this facility.				
BRENTWOOD NORTH NURSING & REHAB	113	50 minutes	14 minutes	1 hour 4 minutes	1 hour 56 minutes
HIGHLAND PARK HLTH CR CENTER	82	40 minutes	5 minutes	45 minutes	1 hour 30 minutes
MANORCARE OF HIGHLAND PARK	85	1 hour 4 minutes	18 minutes	1 hour 22 minutes	2 hours 3 minutes

- Relying on these dimensions is valid if two criteria are met

- . They are correlated with the things that matter
- . You do not encourage “teaching to the test” behavior that leads to higher measured scores but lower quality on the dimensions that really matter
- . The latter is known as *multitasking*; we will return to this later

Nursing Home Compare also reports the first category of quality measures. Here is a listing:

✓	<b>Long-Stay Residents are people in an extended or permanent nursing home stay</b>	
✓	<u>Percent of Residents Whose Need for Help With Daily Activities Has Increased</u>	
✓	<u>Percent of Residents Who Have Moderate to Severe Pain</u>	
✓	<u>Percent of High-Risk Residents Who Have Pressure Sores</u> AND <u>Percent of Low-Risk Residents Who Have Pressure Sores</u>	
✓	<u>Percent of Residents Who Were Physically Restrained</u>	
✓	<u>Percent of Residents Who are More Depressed or Anxious</u>	
✓	<u>Percent of Low-Risk Residents Who Lose Control of Their Bowels or Bladder</u> AND <u>Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder</u>	
✓	<u>Percent of Residents Who Spend Most of Their Time in Bed or in a Chair</u>	
✓	<u>Percent of Residents Whose Ability to Move About in and Around Their Room Got Worse</u>	
✓	<u>Percent of Residents With a Urinary Tract Infection</u>	
✓	<u>Percent of Residents Who Lose Too Much Weight</u>	

Avedis Donabedian proposed a way to classify different measures of quality

- *Outcomes*: the measure we are most interested in. Outcomes include mortality and various measures of morbidity.
- *Credentials*: Schooling, accreditation, affiliations
- *Inputs*: Number and quality of personnel and equipment
- *Process*: Indicators of what, when, and how health services are delivered; e.g., vaccination rates, lengths of stay, tests ordered, etc.

To this, we can also add

- *Volume*: Providers with more experience have superior quality
- The volume/outcome relationship has been well-documented
  - . Q: Can you offer two distinct explanations?
  - . Does it matter to consumers which explanation is correct?

Nursing Home Compare relies on outcomes and inputs, but all of these measures have been used in quality reports

Q: Are some of these measures more important than others?

A recent *NEJM* (2002) study demonstrates an important link between nurse staffing and several outcomes measures

- Main purpose of study was to raise the alarm about the nursing shortage
- But to do so, the authors needed to link inputs to outcomes
- Table 3 in the handout has some of the key results
  - . Coefficient  $< 1$  indicates that more nurses are associated with lower incidence of adverse outcomes

## Quality Report Cards

Until the 1980s, report cards were virtually nonexistent. Patients therefore did not have a full scope of options when shopping for health care providers

- They relied on personal experience, experiences of friends and family, and the recommendations of their physicians
- Brand mattered in a limited way (e.g., Mayo)
- Third party evaluation of provider quality was nonexistent

By early 1980s, researchers and entrepreneurial health care consultants began to measure and report provider outcomes quality

- These report cards can be valuable if two demand-side conditions are met:
  - 1) Patients are unable to discern differences in provider quality
  - 2) Patients read, understand, and act on report card information (or are influenced by others who read, understand and act on them)
- Assuming that these conditions are met, the theory of quality choice tells us that there are three major benefits of reporting outcomes quality
  - 1) Patients can do a better job of finding the best quality providers
  - 2) Patients can find providers who best meet their needs (triage)
  - 3) Providers have incentives to improve their quality

While the theory supporting provider quality reports is self-evident, the practical matter of measuring outcome quality is something else

- There are several critical issues:
  - . How do we use available data to measure outcomes?
  - . If not outcomes, then what else?
  - . How do you assure that these measures are not biased?

Consider the types of data available, on a systematic basis, to all researchers

- *Electronics claims data* are widely available (E.g. CA OSHPD)
  - . Medicare claims data available to anyone for a (steep) price
  - . Some Medicaid data also available
  - . Most states disseminate hospital utilization data for all inpatients (only two states provide MD information)
  - . Insurers and employers have claims data and some have formed consortiums to pool the data
  - . Data include charges but not prices
  - . Data include limited diagnostic and procedure information
- *Electronic medical records* are more difficult to access
  - . Perhaps half of hospitals have them
  - . Much smaller percentage of doctors
  - . Incompatible software, with third parties now emerging to try to get the data to “communicate” by acting as “translators”
  - . Include detailed diagnostic information
  - . Include some outcomes information (in-hospital mortality; diagnoses revealing post-operative infections; post-operative lengths of stay)
  - . It is possible to link Medicare data to Social Security death records to obtain post-discharge mortality for Medicare patients
  - . This leaves out lots of outcomes information, including mortality for non-Medicare patients, and virtually all forms of morbidity
- Despite present data limitations, there are lots of high profile and well-funded efforts to measure and disseminate quality scores

- Here are CABG scores from healthgrades.com (The scores are for in-hospital mortality, in-hospital + 1 month, and in-hospital + 6 months)

**In-hospital Mortality (Survival):** This term refers to a measure of the number of patients that died while still in the hospital\* as compared to the total number of patients having the same procedure in the same hospital. Actual data is compared to predicted data that is determined using a risk-adjustment model.

2006 Ratings			
Coronary Bypass Surgery			
High Volume Hospitals			
	Inpatient Mortality	Inpatient+30	Inpatient+180
<b>+ <u>Swedish Covenant Hospital</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Advocate Ravenswood Medical Center</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Saint Joseph Hospital</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Northwestern Memorial Hospital</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Saints Mary and Elizabeth Medical Center Division St Campus</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>University of Illinois at Chicago Medical Center</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Mercy Hospital and Medical Center</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>Rush University Medical Center</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★ ★ ★
<b>+ <u>University of Chicago Hospitals and Health System</u></b> Chicago, IL	★ ★ ★	★	★ ★ ★
<b>+ <u>Resurrection Medical Center</u></b> Chicago, IL	★ ★ ★	★ ★ ★	★

Q: How do you think University of Chicago would respond to the reportedly low score?

The claim that “my patients are sicker” is the "holy writ" of quality measurement

- As important it is to measure outcomes, it is perhaps more important to do a good job of adjusting for severity
- If the best providers (those who have the best chances of healing any specific patient) tend to treat the sickest patients, then their raw mortality scores may be higher than those for lower quality providers
- With proper severity adjustments, we can estimate the true underlying ability of each provider
  - . Patients would have better expected outcomes if they visit higher-ranking providers (I.e., the rankings are not misleading)
  - . This will enhance patient sensitivity to actual quality differences, encouraging providers to improve their quality

Let's return to report card construction

Computing the "severity-adjusted" mortality score

- There is a consensus methodology for reporting mortality scores
- Researchers compute two numbers – the actual mortality rate and the expected rate. The expected rate is derived from regression
- Either the ratio between the two, or the difference between the two, is a measure of quality

Here is how it is done formally. We need some notation:

$i$  denotes the patient,

$j$  denotes the provider

$death_{ij}$  is a dummy variable that equals 0 if the patient lives and 1 if the patient dies

$S_i$  is a measure (or set of measures) of the severity of patient  $i$ 's illness

$n_j$  is the number of patients treated by provider  $j$



The observed mortality rate ( $OMR$ ) for provider  $j$  is given by

$$OMR_j = \sum(\text{death}_{ij})/n_j$$

Obtaining the expected mortality rate ( $EMR$ ) requires three steps

- 1) Perform a regression:  $\text{death}_{ij} = B_0 + B_1S_i + \varepsilon_i$ . Ideally, this should be estimated using logit, logistic, or probit regression (special regression programs designed to work with dichotomous dependent variables). Ordinary least squares regression (OLS) is adequate for the class project.
- 2) Use the regression coefficients to compute the predicted probability of each person dying – let's call this  $p\text{death}_{ij}$
- 3) For all the patients who visited doctor  $j$ , compute  $EMR_j = \sum(p\text{death}_{ij})/N_j$

This gives us the  $OMR_j$  and the  $EMR_j$

- But this does not tell us if the difference is statistically significant

We need some to compute some additional statistics

- We have a measure of  $EMR_j$  for each provider  $j$ , but it is only an estimate
  - . It is estimated imprecisely, because our model predicting  $\text{death}_{ij}$  is imprecise
  - . If we are determine if  $OMR_j$  is statistically different from  $EMR_j$ , we need to know the confidence interval around  $EMR_j$
- To compute the confidence interval, we must calculate the standard deviation of  $EMR_j$ , or  $\sigma_{EMR_j}$ . This is not easy to do without additional training in stats and some good software. We will not drill this deep.
- We can use  $\sigma_{EMR_j}$  to calculate the confidence interval around  $EMR_j$ 
  - . Use the rule of thumb that the 95% confidence interval around any estimate is  $\pm 2$  times the standard deviation of that estimate
  - . Thus, our confidence interval is  $EMR_j \pm 2 \cdot \sigma_{EMR_j}$

You now know, in theory, how to compute  $OMR_j$ ,  $EMR_j$ , and with some statistical help you could also compute  $EMR_j \pm 2\sigma_{EMR_j}$

Here is how to interpret the results:

- If  $OMR_j > EMR_j$ , the provider had more deaths than predicted
- If  $OMR_j > EMR_j + 2\sigma_{EMR_j}$ , then the excessive death rate is statistically significant
- Your project requires you to compute  $OMR_j$  and  $EMR_j$ . You can do this in K-Stat.

Note that some report cards report  $RAMR_j = OMR_{average} \cdot (OMR_j/EMR_j)$

- Confidence intervals are computed using the same  $\sigma_{EMR_j}$ ; however, the formula is a bit more complex.
- E.g. New York's ratings (overheads)

Report card critics observe that the regressions used to estimate  $death_{ij}$  are very imprecise, with relatively low  $R^2$ .

This necessarily translates into large  $\sigma_{EMR_j}$  and large confidence intervals

- E.g., In the inaugural Hospital Compare report card launched by CMS last year, only 3 percent of hospitals had statistically significant deviations from expected mortality
- Q: Does that render the report card useless?
- In any event, superior regression models and data would make report cards more powerful

Unfortunately, our models predicting death and other outcomes are rather poor, both for statistical and data reasons

- Statistics: It is difficult to develop a good predictive model when almost all the outcomes are 0 (or if almost all the outcomes are 1)
  - . This implies that if mortality is the outcome, you can only study conditions for which the mortality rate is fairly high (at least a few percent).
  - . This drastically limits the scope of diseases/procedures for which we can report reliable mortality scores (CABG is the favorite)
  
- Data: Severity adjusters in administrative claims data are very limited
  - . We usually have age, sex, and secondary diagnoses
  - . The diagnoses are not consistently reported across providers. Many providers are now trained to report those codes required to make their patients appear sicker!
  - . Historical utilization data can be used as a risk adjuster but might instead reflect poor quality in the past. Using it as a risk adjuster gets the provider “off the hook”
  
- Exceptions: Some analysts, such as in NYS, have access to medical records that permit use of more sophisticated risk adjusters

Perhaps because of these limitations, many report cards go beyond reporting outcomes. Consider this additional information in the CMS Hospital Compare

<b>Heart Attack Care Quality Measures</b> (some of the recommended care given to patients if appropriate*) <input type="checkbox"/> <b>Select All Heart Attack Care Measures</b>	<b>Brief Explanation of Treatment</b>
<input type="checkbox"/> Percent of Patients Given ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	ACE (angiotensin converting enzyme) inhibitors and ARBs (angiotensin receptor blockers) are medicines used to treat heart attacks, heart failure, or a decreased function of the heart.
<input type="checkbox"/> Percent of Patients Given Aspirin at Arrival	Aspirin can help keep blood clots from forming and dissolve blood clots that can cause heart attacks.
<input type="checkbox"/> Percent of Patients Given Aspirin at Discharge	Taking aspirin may help prevent further heart attacks.
<input type="checkbox"/> Percent of Patients Given Beta Blocker at Arrival	Beta blockers are a type of medicine used to lower blood pressure, treat chest pain (angina) and heart failure, and to help prevent a heart attack.
<input type="checkbox"/> Percent of Patients Given Beta Blocker at Discharge	Beta blockers are a type of medicine used to lower blood pressure, treat chest pain (angina) and heart failure, and to help prevent a heart attack.
<input type="checkbox"/> Percent of Patients Given PCI Within 120 Minutes Of Arrival	The procedures called Percutaneous Coronary Interventions (PCI) are among those that are the most effective for opening blocked blood vessels that cause heart attacks. Doctors may perform PCI, or give medicine to open the blockage, and in some cases, may do both.
<input type="checkbox"/> Percent of Patients Given Smoking Cessation Advice/Counseling	Smoking is linked to heart attacks. Quitting may help prevent another heart attack.
<input type="checkbox"/> Percent of Patients Given Thrombolytic Medication Within 30 Minutes Of Arrival	Blood clots can cause heart attacks. Doctors may give this medicine, or perform a procedure to open the blockage, and in some cases, may do both.

## Here is a complete set of quality measures from the private sector Leapfrog Group

Search Results: **Zip:** 60035 **Radius:** 10 Miles

Below are the results of your search. Click on the "leaps" and the circles for more details.

Click to Compare	Hospital Name	City	Leap1	Leap2	Leap3					Leap4	Results Submitted	
			CPOE	ICU	High Risk Treatments					Safe Practices Score		
	<a href="#">Glenbrook Hospital</a>	Glenview, IL										6/28/2006
	<a href="#">Highland Park Hospital</a>	Highland Park, IL										6/28/2006
	<a href="#">Lake Forest Hospital</a>	Lake Forest, IL										6/30/2006

Leapfrog asks hospitals if they adhere to the following four quality and safety practices. Significant scientific evidence shows that these practices reduce unnecessary deaths and injuries.

1. **CPOE** – choose a hospital that requires its staff to use computers to order medications, tests and procedures. This is called a Computerized Physician Order Entry system (CPOE). [More information on Computerized Physician Order Entry.](#)
2. **ICU Staffing** – choose a hospital with an intensive care unit (ICU) that is staffed by doctors and other caregivers who have special training in critical care. These are doctors are called 'intensivists'. [More information on ICU Staffing.](#)
3. **High Risk Treatments** – select a hospital with lots of experience and the best results for specific procedures, surgeries or conditions. This is known as Evidence Based Hospital Referral (High risk procedures). [More information on Evidence-Based Hospital Referral.](#)
4. **Leapfrog Safe Practices Score** – select a hospital that has a high Leapfrog Safe Practices Score. This means it has put in place 27 procedures to reduce preventable medical mistakes. More information on the [Leapfrog Safe Practices Score.](#)

Note that the “high risk treatment” scores are volume based. The Niagara Healthcare Quality Coalition also relies heavily on volume based report cards

*In-class discussion question:*

Kellogg alum Bruce Boissonault, Director of the Niagara Health Quality Coalition, believes that the NHQC report cards are the best in the nation. What do you think?

In what ways are the NHQC, Healthgrades, and Leapfrog reports similar? How do they differ?

In what ways are the NHQC reports more useful? In what ways are they less useful? What more would you want to know before comparing different report cards or choosing a hospital based on one or more report cards?

Most patients remain unaware of the mortality reports

- The best known report card remains the US News and World Reports, which ranks the top hospitals only

- . These rankings are based as much on reputation among providers than anything else

- . If all patients tried to be treated at the ranked hospitals, 90 percent would have to be turned away

Nor are providers are not eager to promote the existence of report cards (unless they score high)

- Most physicians surveyed do not believe that their patients should see report card scores

This raises an important empirical question: Do report cards move market share?

Q: Why is it important that they do? If they don't, then are report cards worthless?

Q: How would you set about testing this? (Defer)

*More on Report Cards: The Experiences of NY and PA*

New York issued its first cardiovascular surgery report cards in 1991; PA in 1993

- NY augmented severity adjusters in claims data with information pulled directly from chart review
- More on NY later

PA's innovation was based in economics rather statistics

- Most published report cards include multiple dimensions of quality
  - . This might be too much to process.
  - . Might be difficult for consumers to construct an overall ranking
- PA asked economists at U of Pennsylvania to create a meaningful composite score
  - . Composite scores require weighting individual components
  - . Q: How do you construct the weights?
  - . A: *Revealed preference* methods
- PA scored hospitals on three dimensions: mortality, major morbidities, and complications
  - . Researchers asked individuals to rank hospitals with various combinations of the different outcomes
  - . E.g. Hospital A has (Mortality = .02, Complications = .40)  
Hospital B has (Mortality = .03, Complications = .30)  
Which do you prefer?
  - . If patient is indifferent, then mortality should get ten times as much weight as complications
- The weights they obtained were as follows:  
$$\text{Score} = .46(\text{Mortality}) + .29(\text{Major Morbidity}) + .25(\text{Complications})$$

Reactions to the NY and PA report cards were mixed

- Hospitals invoked the holy writ
- But many hospitals attempted to boost quality
  - . Hospitals dismissed problematic physicians, replaced department heads, retrained staff, etc.
  - . Poor performers were most likely to take action
  - . Hospital boards were catalysts for change
- Report cards seem to produce positive statistical changes as well
  - . Risk-adjusted cardiovascular surgery death rates fell relative to national trends
  - . At same time, reported severity of patients rose

Q: Recall how report card scores are constructed. Does the combination of falling risk-adjusted mortality and rising reported severity necessarily imply that quality is improving?

Some analysts worried that providers would game the system to improve their scores

- “Upcode” patients to increase reported severity
- Shun patients who are sicker than the “regression model” would predict
- Treat patients who are healthier than predicted
- Dranove et al. (2003) explore the latter two behaviors



## Overview of Dranove et al. study

- NY and PA CABG report cards measure outcomes for patients receiving surgery but ignore patients who receive alternative therapy
- This can create a statistical problem, as the following example shows
- Suppose there are equal numbers of “High risk” and “Low risk” patients
- Suppose there are two types of treatments (Surgical and Medical)

$R(i,j)$  = mortality risk;  $i$  = patient type,  $j$  = treatment type

$$r(H,S) = .1 \quad r(H,M) = .14 \quad r(L,S) = .04 \quad r(L,M) = .02$$

Use these data to compute mortality rates for different rates of surgical intervention:

*Regime 1:* All type H receive S; All type L receive M

Surgical mortality rate = .10

Overall mortality rate = .06

*Regime 2:* 75% of type H receive S; 75% of type L receive M

Surgical mortality rate = .085

Overall mortality rate = .0675

Note: S population has lower mortality rate in regime 2. But the "wrong" people got surgery, so the overall mortality rate falls

Conclusion: One cannot evaluate surgical report cards by examining outcomes for surgical patients

Dranove et al. set about determining whether this kind of gaming took place in the first two years after the introduction of report cards (thus, they do not measure the *long term* effects of report cards)

- We focus on Medicare patients with AMI – they all end up at a hospital and their medical data can be linked to Social Security death records
- We use prior year's expenditures and/or prior year's admission as their measures of severity
  - . These measures do predict outcomes and are not used by NYS or PA to predict outcomes
  - . Thus, they are good ways to measure severity that are likely immune from upcoding
- Table 1 presents important preliminary evidence
  - . Relative to other states, NY and PA AMI patients who receive surgery were getting “healthier” on measures not used for risk adjustment (lower prior year expenses)
- Tables 4 and 5 answer the following questions
  - . How did introduction of report cards affect various outcomes?
  - . Is effect larger for the “sickest” patients? (Here, sicker means “prior year admission”)
- Here are the answers:
  - . Introduction of report cards seems to increase the likelihood of receiving CABG (within one year of AMI) for “healthier” patients
  - . Healthier patients were less likely to get PTCA or catheterization
  - . All three interventions were delayed, for healthy and (to a lesser extent) sick patients
  - . Costs went up for all patients, especially sicker patients
  - . Readmissions were higher for sicker patients
  - . Overall mortality rate was higher for sicker patients

In another important study, Werner examines how the NY report card affected surgical rates for blacks and Hispanics

- Race is a predictor of outcomes, but NY does not use race in its adjustment
- Doctors may use race when deciding whether to go ahead with a surgery
- Even before the report cards, there were racial disparities in surgical rates
  - . There was a 1.6% disparity in white versus black intervention rates in NY, 3% in the rest of the nation
  - . After the report cards, the disparity in NY rose to 4% and in the rest of the nation rose to 3.3%
  - . Trends for white versus Hispanic were even starker

### **Patient Responses to Report Cards**

Many critics of report cards point out that few patients seem to know about them

- Providers' intrinsic motivation to get better can only take quality so far
- Without a demand response, report card effectiveness will be limited

How many patients need to know?

- Word of mouth is a powerful demand driver
- MDs can base their referral decisions on report cards, thereby eliminating need for patients to stay abreast of quality information

Empirical question: Do report cards move market share?

- Early evidence was discouraging
  - . High ranking hospitals did not seem to gain market share
  - . The Mayo conundrum: What if most high ranking hospitals were already highly regarded?
  
- Relevant question: Do report cards move market share when they contain “news” about quality
  - . “News”: Report card rankings do not conform to prior rankings of quality based on general reputation
  - . I.e., we would not expect a high ranking to benefit Mayo, but it might help “Community General Hospital”
  
- Dranove and Sfekas (2008) examine demand response to news
  - . Examine NYS hospitals before and after initial report cards
  - . We measure news by comparing report card rankings with quality as measured from pre-report card demand models
  - . We show that news does move market share. Hospitals with the best “news” saw demand increase by 8%; a similar decline was experienced for hospitals with bad “news”
  - . A \$1 million annual impact on the bottom line

## Pay for Performance

Whether it is out of concern that patients are not paying enough attention to report cards, or concern about the statistics of report cards, payers are taking a more direct approach to quality improvement by paying for performance

- Many insurers offer explicit rewards to providers who meet certain measurable objectives.
- For example, the Leapfrog Group has set standards that many payers now use to base payment rates and bonuses
- Blue Cross of Illinois and UHC of Illinois are two of many private insurers with direct rewards to providers who meet performance objectives
  - . Time to schedule appointments
  - . Smoking cessation advice
  - . Retinal exams for diabetics

The theory of pay for performance seems rather straightforward

- If you pay someone more for doing X, they will do more of X

But there are pitfalls:

- Paying for X may lead to less of Y (“multitasking”)
- Provider may not be responsible for providing X
- It may be costly to measure X (record keeping costs)

## Multitasking

Most work, especially professional services, is complex

- Sometimes, performing better on one dimension can enhance performance on all dimensions
  - E.g., Better surgical skill can lead to faster healing and fewer complications
- Economists say that these dimensions display *complementarities*

But sellers may have limited resources to apply to their job – limited time, limited personnel, and limited cash

- Improving their performance on one dimension (e.g., spending more time with the patient) may necessarily require reducing performance on another dimension (spending time making a diagnosis)
- In these cases, the various dimensions of performance are substitutes

If a purchaser cares about dimensions X and Y but can only measure/pay for X, then:

- The provider will provide more of X and more of Y if they are complements
- The provider will provide more of X and less of Y if they are substitutes

Summarizing:

	Pay for X and:
X and Y are complements	Get more of X and more of Y
X and Y are substitutes	Get more of X and less of Y

Example: Hospital tradeoffs between CPOE and nurse staffing

In the past year, two studies have looked at multitasking and provider performance evaluation. One, by Kellogg PhD student Susan Lu, examines nursing home responses to nursing home compare

- Recall that nursing home compare publishes lots of process and input measures of quality
- One statistic collected by CMS but not reported is the severity of patient-reported problems.
- Lu reports that published scores are improving, but that patient-reported problems are becoming more severe. Is this evidence of multitasking?

Another study finds that doctors are providing more preventive services from a list of those included in P4P, while cutting back on those not included

Q: What should payers do?

Here are some additional issues that report card advocates are wrestling with

1) What data to use for risk adjustment?

The choice is between administrative claims data and medical records

Administrative claims data:

Pros: Easy to obtain; can be linked across providers

Cons: Limited clinical information, thereby limiting the predictive power of the model; limited outcomes data

Medical record data:

Pros: Detailed clinical information enhancing predictive power and outcomes measurement

Cons: Hard to obtain; hard to link across providers; diagnostic information can be manipulated via upcoding

2) How many games can providers play?

Upcoding is one way to game the system

Patient selection is potentially far more harmful

- Select patients who afford the best opportunity for obtaining high scores
- This applies to both report cards and pay for performance
- The result can be a net negative for the system, as the wrong patients get care

3) Can you measure what matters?

- Mortality is the easiest outcome to measure. But most costs are associated with chronic conditions

- Can we collect more and better outcomes data?

. E.g., SF-10 for Children

. [http://www.caremark.com/portal/asset/CP\\_SF-10.pdf#search=%22sf-10%20for%20children%22](http://www.caremark.com/portal/asset/CP_SF-10.pdf#search=%22sf-10%20for%20children%22)

. Includes questions like “During the past 4 weeks, has your child been limited in bending, lifting, or stooping due to health problems?”

- Q: Can we link it to providers? Who will pay for this?

4) Report card “confusion”

- Report cards are being mass-produced

- In many states, you have state report cards, employer report cards, and provider-sponsored report cards. Add to this USNews, Healthgrades, etc., and we have report card overload

- Every provider can find a report card that puts them on top



- Some report card advocates claim that providers are intentionally releasing report cards of questionable validity so as to confuse consumers and limit the effectiveness of valid reports

#### 4) Is Anyone Listening?

Recall the twin benefits of quality evaluation:

- (a) patients more likely to visit higher quality providers, and
- (b) added incentives for providers to boost quality

It seems that (a) is necessary for (b)

- Why would providers boost quality if patients don't care?
- The fact is that when hospitals get bad scores, board members panic
- Perhaps report cards can encourage the bad to become average, but what will motivate the average to become great?

#### **MCO quality**

MCOs make many decisions that can affect quality

- Network design and restrictions on referrals
- Enforcement of UR
- Compensation incentives

On all of these dimensions, MCOs have much in common

- They use nearly identical networks and often rely on the same UR service agencies
- Within MCO "type", they have fairly similar schemes for paying providers

Subtle differences in any of the MCO "levers" could have a profound impact on quality

- As quality model reminds us, if quality is unmeasured, MCOs could skimp on quality
- Properly measured, MCOs will have incentive to boost quality

Serious efforts to measure MCO quality began in early 1990s

- A consortium of employers and HMOs formed NCQA
- They developed HEDIS, a multidimensional report on HMO quality

. Here is a link to a representative HEDIS report

**<http://www.health.state.mn.us/divs/hpsc/mcs/hedis/blue05.pdf>**

- HEDIS reports contain hundreds of "quality" measures
  - . "Effectiveness of care" includes mammogram rate in eligible women, childhood immunization status, beta blocker prescription following heart attack)
  - . "Use of services" includes well child visits, hospital discharges, outpatient surgery rates, and rates of specific procedures such as hip replacement and CABG
  - . "Access/Availability" includes percent of members who have a preventive visit, percent of pregnant women who have "timely" prenatal care visit
  - . "Plan stability" includes turnover rate in contracting providers, and years and business and size of plan

One final question remains: Do HEDIS scores affect plan choice?

- Many employers use HEDIS or related measures
- There are many surveys suggesting limited consumer knowledge and use of HEDIS

Recent research by University of Michigan studies how report cards affect consumer choice of plan

- Scanlon et al. take advantage of a “natural experiment”, whereby GM requires its plans to submit report cards, and then GM summarized the results for its employees (using *Consumer Reports*-style ratings)
- They then regress enrollments on the number of “superior counts” and “below average counts” on the various dimensions of quality
- Interestingly, superior counts do not affect enrollments
- However, below average counts do reduce enrollments

Dafny and I have recently examined how reporting of HEDIS scores to Medicare HMO enrollees affected enrollments

- Medicare & You publication began including selected scores in 2000
- Not clear if seniors read the publication or if they would respond
- Key findings:
  - . High quality plans were gaining share even before report cards released
  - . Report cards had an effect equal to about 7 years of market share movement prior to report cards
  - . Only consumer perception based measures of quality (e.g., “Best Possible Care”) mattered.