MEASURING PROVIDER PERFORMANCE: THE FOUNDATION OF PURCHASER STRATEGIES TO REFORM AMERICA'S HEALTH CARE SYSTEM



Measuring Provider Performance

- ▶ Who constructs measures?
- ▶ Common measurement issues
- Measuring quality
- Measuring cost/efficiency
- Measuring overuse



Who Constructs Measures?

- ▶ Health plans
 - Responding to employer demands for quality and cost information to share with their enrollees to inform choice of providers.
 - Used internally in reimbursement negotiations, setting of reimbursement rates, and constructing tiered networks.
- Community/purchaser coalitions
- Used in public reports of provider performance.
- Independent, non-profit organizations
 - Generate national consensus measures that can be used by health plans and community coalitions.
- Physician specialty and hospital organizations
 - Generate measures acceptable to providers with intention that the measures be used by health plans and community coalitions.

PubH 7564 - Online 2013

| Figure 3.7 | |
|---|--|
| Use of National Quality Forum–Endorsed Measures in Measurement Instances (n = 70 organizations) Quality improvement (n = 75) | |
| Payment (n = 36) Use #All endorsed | |
| Public reporting (n = 81) # Some endorsed Accreditation, certification, credentialing, and licensure (n = 16) Not reported | |
| 0% 10% 20% 30% 40% 50% 60% 70% 80% 90%100% Instances reported (One measure endorsement status per setting of care per measure use per organization, n = 208) | |
| NOTE: The number to the right of each measure use shows the number of discrete organization-setting instances where organizations reported using measurer for that purpose, sale minutes? | |
| Damberg, C.L., Sorbero, M.E., Lovejoy, S.L. et al. An Evaluation of the Use of Performance Measures in Health Care. Technical Report. Santa Monica, CA. RAND Corporation, 2011. | |
| 4 ubH 7564 - Online 2013 | |
| | |
| | |
| Considerations in Selecting Measures* | |
| | |
| Relevance Be relevant to different groups (consumers, clinicians) Pertain to significant health problem or condition | |
| Vary across providers Have some potential for improvement | |
| – Be amenable to change by providers | |
| *http://www.ncqa.org/tabid/425/Default/aspx | |
| | |
| bH 7564 - Online 2013 | |
| | |
| | |
| Considerations in Selecting Measures* | |
| ➤ Scientific Soundness | |
| Supported by clinical evidence (if measure of quality) Produce the same results when repeated in same | |
| population and setting – Makes sense logically and clinically | |
| – Is amenable to risk adjustment | |
| *http://www.ncqa.org/tabid/425/Default/aspx | |
| 6 | |

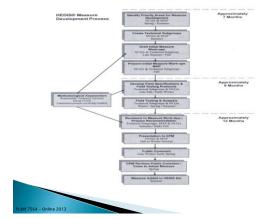
Considerations in Selecting Measures*

▶ Feasibility

- Data are available for constructing measure
- Clear specifications for data sources and for methods of data collection and reporting
- Data collection/provision does not impose unreasonable burden on providers
- Data collection does not violate confidentiality standards
- Measure can be audited to prevent manipulation by providers

*http://www.ncqa.org/tabid/425/Default/aspx

PubH 7564 - Online 2013





| | _ | |
|---|----|--|
| Measure Concepts In Development Measure Concepts on Hold (data source) | | |
| Pediatric Preventive Care High Tech Diagnostic Imaging Total Cost of Care 30 Day All-Cause Readmission Care Coordination [Health Care Homes Only] | | |
| Reasures in Pilot/ Implementation New Measures for Submission | - | |
| Total Knee Replacement Maternity Care- C-Section | | |
| Lumbar Spine Surgery Colonoscopy Quality | | |
| Alinnesota Community Measurement. T <i>he Measurement Minute</i> , August 2012 | _ | |
| | | |
| I – Online 2013 | 10 | |
| | | |
| | | |
| | | |
| | | |
| Common Measurement Issues | | |
| | | |
| ▶ Attribution | | |
| | | |
| ► Composite vs. specific | | |
| ▶ Risk Adjustment | | |
| Validity/Reliability (small numbers problem) | | |
| valuity/Renability (smail numbers problem) | | |
| | 11 | |
| - Online 2013 | | |
| | | |
| | | |
| | | |
| | | |
| Attribution | | |
| | | |
| Problem is more important for measurement of | | |
| physician performance, as opposed to hospital performance. | | |
| Attribution becomes an issue when patient sees | | |
| multiple providers during measurement time | | |
| period, as is often the case for patients with chronic conditions. | | |
| - Whose quality or cost performance is being measured? | | |
| | | |
| | | |
| | 12 | |

| Attribution | | |
|---|----|--|
| Options for attribution: | | |
| Assign responsibility for quality or cost outcomes to physician with most visits from patient during measurement period. | | |
| Assign responsibility to physician who accounted for largest share of costs (for cost measures) | | |
| Assign responsibility to primary care physician Assign measure to all providers seen by patient during measurement period. | | |
| | | |
| uBH 7564 - Online 2013 | 13 | |
| | | |
| | | |
| Attribution | | |
| How to attribute people who don't use care in | | |
| measurement period. In analysis of physician cost profiling, RAND | | |
| concluded that attribution rule (tested 12 variations) strongly affected cost category physician | | |
| was placed in — using different rules resulted in reassignment of up to 61 percent of physicians | | |
| (RAND Research Highlights: Is Physician Cost Profiling Ready for Prime Time?) | | |
| ubh 7564 - Online 2013 | 14 | |
| | | |
| | | |
| Composite vs. Specific Measures | | |
| As number of performance measures grow, it | | |
| becomes more confusing for purchasers and payers to make judgments about high versus low | | |
| performing providers. – Providers may be high performers on some measures and low on others. | | |
| Relative importance of different measures can be hard for consumers to assess. | | |
| Consumers may be confused, not use information in making decisions. | | |
| ubl 7564 - Online 2013 | 15 | |

Composite vs. Specific Measures ▶ Composite, summary, aggregate, or "all or nothing" measures collapse a large number of measures into a single, or smaller number, of measures. – Easier for consumers to reach judgments about - but involve value judgments (that may not be clear) in collapsing measures. **Creating Composite Measures** ▶ Denominator-based weighted measures - Simple average of all indicators - Weighted average, based on number of eligible patients - "All or nothing" ▶ Presentation - Composite measure only - Composite measures with measures on specific indicators - Ranking of providers – Star rankings (e.g. quartile in which provider's score

| Risk Adjustment | | |
|--|----|--|
| Problem: Patients are not randomly distributed across providers; some providers may attract more severely ill patients than others, even within a specific condition group (e.g. diabetes). Performance measures may indicate that a provider is more costly, or of | | |
| lower quality, if the provider attracts sicker patients and/or patients with other characteristics that make them more difficult or costly to treat. | | |
| Could discourage providers from accepting these patients. Could unfairly rank some providers higher than others, or pay them more, simply because they treated less complex, or less "compliant," | | |
| patients. Goal of "risk-adjustment": Take into account inherent differences in patients when measuring performance. | | |
| In quality measurement, more important for patient outcome than treatment process measures. | | |
| bH 7564 - Online 2013 | 19 | |
| | | |
| | | |
| D' L A l' | | |
| Risk Adjustment | | |
| • Approaches to problem: | | |
| Group patients into "cells" with separate scores computed for each cell (e.g. gender); sometimes called "segmentation". | | |
| Use statistical methods to "control for" differences in patient characteristics; large number of statistical | | |
| approaches are available (some are proprietary). | | |
| | | |
| bH 7564 - Online 2013 | 20 | |
| | | |
| | | |
| Risk Adjustment | | |
| Factors included in risk-adjustment models: Demographic information | | |
| Self-reported health status Administrative data (medical and pharmacy claims) indicating patient condition, diagnosis, past expenditures | | |
| Demographic information: easy to collect, accurate, but explains relatively little variation. Self-reported health status: expensive to collect as it requires | | |
| surveys, adds slightly to explanatory power of demographic information. | | |
| Administrative data: adds considerably to predictive power, especially for expenditures; may create undesirable incentives for providers; diagnosis information not always complete | | |
| bH 7564 - Online 2013 | 21 | |
| | | |

Validity/Reliability

- Question: How accurately does measure reflect actual performance?
 - In any measure, there is likely to be some error.
 - One criterion for measure selection is the reliability of the measure.
- Problem of small numbers
 - Generally accepted that a minimum of 30 patients is needed per $physician\ or\ physician\ practice\ to\ construct\ reliable\ measures\ of$ performance
 - The number may need to be much higher when there is large variation in performance measure

 — It is unusual for physicians to have enough patients with a specific
 - condition enrolled in any single plan to construct reliable measures for plan enrollees.

 - The even for relatively common conditions like diabetes

 When the number of patients for whom performance is measured (assigned to a single physician) is low, the risk of misclassifying physicians as "high" or "low" performers is great.

| ubH 7564 - Online 2013 | |
|------------------------|--|
| | |

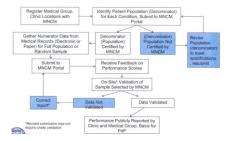
Measuring Quality

- ▶ Process versus outcome measures
- Process measures
 Often can be calculated with claims data, so are relatively cheap to

 - Can address preventive or acute care
 Actual measures are typically based on HEDIS measures, so are familiar to health plans and providers.
- Physicians dispute their accuracy (don't trust accuracy of claims)
 Outcomes measures
- - Typically cannot be constructed using claims data, so require chart
 - review or abstraction of data from electronic medical records.

 Closer to ultimate goal of measuring health of patients of different
 - providers.
 - Providers worry that some outcomes are influenced by patient behaviors as much as the medical treatment they provide; don't want the "blame" for patient's bad lifestyle choices.

Process Flow



| What to measure? - Induction and a survice (a.g. number of visits) delivered - General Control Contr | Measuring Cost/Efficiency | |
|--|--|----|
| Measuring Efficiency in California's Integrated Healthcare Association Program * Program Overview - Integram Ove | Productivity: amount of service (e.g. number of visits) delivered with a given set of inputs (clinical staff/space) Or amount of inputs used to deliver a given service - Cost of Service: cost of providing a given unit of service (e.g. cost per hospital stay) - Cost of Episode: cost for all services used in an episode of care - Cost of Covered Life: cost of delivering all services needed in a given time period Assumption: Costs measured for a given level of quality. Feasibility: Payments typically measured rather than costs Growing interest in developing measures of "value": Quality | 25 |
| Measuring Efficiency in California's Integrated Health Commission State of the | Definition of the control of extraction. Definition matter is an invariant to be precised on a first production of a house invalidation of the animal entire definition of the production of the animal entire definition of the production of the animal entire definition of the animal ent | |
| Measuring Efficiency in California's Integrated Healthcare Association Program* Program Overview Intituted P4P program in 2003 involving pooling of data from 6 California health plans. Built plans. | Allowed amounts: The most annount of control of the proper to targe a plan amount of control of the proper to targe a plan amount of control of the proper to targe a plan amount of control of the proper to targe a plan amount of the proper to target a plan amount of the proper to target a plan amount of th | |
| Measuring Efficiency in California's Integrated Healthcare Association Program* Program Overview Initiated P4P program in 2003 involving pooling of data from 6 California health plans. 6.7 million enrollees, 230 physician groups. Rewarded providers for performance on a single set of quality measures Decided to add efficiency data in 2005. Mechanics In 2006, claims data for 5 million HMO enrollees were run through a commercial "grouper" program, with each claim assigned to one of \$70 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. | sensor down is throughed beginners who down leaves here there there was not been been been sensor the sensor when the sensor was not been to be the sensor when the sensor was not been to been to be the sensor was not been to be the sensor was not been to be the sensor of the sensor was not be the sensor of the sensor was not been to be the sensor of the sensor was not been to be the sensor was not to be the sensor was not to be the sensor of the sensor was not to be the sensor of the s | |
| Measuring Efficiency in California's Integrated Healthcare Association Program* Program Overview Initiated P4P program in 2003 involving pooling of data from 6 California health plans. 6.7 million enrollees, 230 physician groups. Rewarded providers for performance on a single set of quality measures Decided to add efficiency data in 2005. Mechanics In 2006, claims data for 5 million HMO enrollees were run through a commercial "grouper" program, with each claim assigned to one of 570 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs: average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. *Source Robbrook, Williams and Yanaghlara, Health Affaira, September/Octobec 2009. | Robert Wood Johnson Foundation, Counting Change: Measuring Health Care Prices, Costs, and Spending, March, 2012 | |
| Healthcare Association Program* Program Overview Initiated P4P program in 2003 involving pooling of data from 6 California health plans. 6.7 million enrollees, 230 physician groups. Rewarded providers for performance on a single set of quality measures Decided to add efficiency data in 2005. Mechanics In 2006, claims data for 5 million HM0 enrollees were run through a commercial "grouper" program, with each claim assigned to one of 570 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs: average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. *Source: Robinson, Williams and Yanagihara, Health Affiairs, September/Octobec 2009. | | |
| Program Overview Initiated P4P program in 2003 involving pooling of data from 6 California health plans. 6.7 million enrollees, 230 physician groups. Rewarded providers for performance on a single set of quality measures Decided to add efficiency data in 2005. Mechanics In 2006, claims data for 5 million HMO enrollees were run through a commercial "grouper" program, with each claim assigned to one of 570 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs: average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. *Source: Robinson, Williams and Yanagihara, Health-Affairs, September/October, 2009. | | |
| Nechanics In 2006, claims data for 5 million HMO enrollees were run through a commercial "grouper" program, with each claim assigned to one of 570 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs: average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. *Source: Robinson, Williams and Yanagihara, Health Affairs, September/October, 2009. | Program Overview Initiated P4P program in 2003 involving pooling of data from 6 California health plans. 6.7 million enrollees, 230 physician groups. Rewarded providers for performance on a single set of quality measures | |
| | Mechanics In 2006, claims data for 5 million HMO enrollees were run through a commercial "grouper" program, with each claim assigned to one of 570 episode types using primary and secondary diagnoses. Efficiency to be measured at medical group level as ratio of observed-to-expected costs for each episode and aggregated across episodes. Expected episode costs: average cost of each episode type across all physician organizations, adjusting for level of severity (stage of disease) and general health status/comorbidities. | |
| | *Source: Robinson, Williams and Yanagihara, Health Affairs, September/October, 2009. | 9. |

| Measuring Efficiency in California's Integrated Healthcare Association Program* | |
|--|---|
| | |
| Problems Encountered | |
| Small Numbers: – Even though measures were at medical group level and were constructed using data from 6.5 | |
| million individuals, only for 75 (of 570) | |
| episodes did the majority of physician organizations have 30 or more patients per | |
| year, and many of these were for minor conditions. | |
| - Some common conditions (e.g. type 2 diabetes) | |
| did reach this threshold at the group level but | |
| not if measured at individual physician level. | |
| *Source: Robinson, Williams and Yanagihara, Health Affairs, September/October, 2009. | |
| 28 | |
| 28 28 29 29 29 29 29 29 29 29 29 29 29 29 29 | |
| | |
| | |
| | |
| | |
| Measuring Efficiency in California's Integrated | |
| Healthcare Association Program* | |
| | |
| ▶ Problems Encountered | - |
| Data Completeness: – claims data were not complete especially where groups were paid by capitation. | |
| in addition to missing claims, coding was not complete enough for requirements of | |
| episode grouper; secondary diagnoses | |
| were lacking. – Use of actual versus standardized prices | |
| Actual prices paid by insurers varied widely and reflected primarily relative negotiating power. | |
| Medical groups objected to having their efficiency measures influenced by this. | |
| Standardized prices were used, which resulted in a measure of | |
| utilization rather than costs. *Source: Robinson, Williams and Yanagihara, Health Affairs, September/October, 2009. | |
| | |
| 29 PubH 7564 - Online 2013 | |
| | |
| | |
| | |
| | |
| 700 | |
| Measuring Efficiency in California's Integrated | |
| Healthcare Association Program* | |
| | |
| Modifications | |
| Alternative measures used (examples): Generic prescriptions as a percent of all prescriptions (where | |
| generics were available) | |
| Readmissions as a percent of total hospital admissions. Emergency room visits per 1,000 enrollees. | |
| - 0 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 | |
| | |
| | |
| *Source: Robinson, Williams and Yanagihara, Health Affairs, September/October, 2009. | |
| | |
| 30 30 30 30 30 30 30 30 30 30 30 30 30 3 | |

| Mea | suring Overuse | | |
|---|--|---|--|
| Traditional quality meas | sures have focused on documenting underuse of for instance, percent of physician patients | | |
| | service, with smaller percentage indicating worse | | |
| Overuse: "The use of a which potential harms | " of services has become a focus for measurement service that is unlikely to improve patient outcomes or for sexceed likely benefits." s that physician reduce use of particular procedure or | | |
| Reaching professional Examples include som tumors. | consensus on measures of overuse is difficult the types of CT scans, chemotherapy for some types of | | |
| | ne potential to be "win-win" – improving quality or the health care system but a revenue loss for service payment. | | |
| °Mathias and | Baker, 'Developing Quality Measures to Address Overuse, JAMA, May 8, 2013, p. 1897. | | |
| | | | |
| | | | |
| Table. Potential Unintended Consequences of Ox | veruse Measurement | | |
| Potential Unintended Consequences of Overuse Measu Undoruse of the service when it is actually indicated | urement Example Measuring overuse of bone scenning for staging low-risk patients may unintentionally lead to | | |
| Underuse of other related services | underuse in higher-risk patients for whom a bone scan is indicated. Measuring overuse of cervical cancer screening in low-risk women may unintentionally lead to | | |
| Patient selection | Measuring overses of spine imaging may unintentionally lead primary care physicians to avoid query | | |
| Care location shift | for patients with low back pain by retening them all to specialists Measuring over the of imparts in observance offices may unintentionally level to observe operations. | | |
| Increesing use of alternate tests or treatments | patients to the emergency room for imaging Moscuring oversee of lumbar spine radioprophs may unintentionally lead to increased use of lumbar. | | |
| Damage to the parient-physician relationship | spine magnetic resonance imaging. Mississing comuse of antibiotics for bronchite may unintendionally damage a physician's relationship with a patient because the physician did not order the treatment the patient dealed. | | |
| Cinician dissatisfaction with quality measurement | Excessive measurement burden (so, additional documentation to confirm that service use is not | | |
| Adverse public health effects | outroue) may lead to clinician dissettatiction with quality measurement on the force insic, gaming the system, at both, to provide partners on the mediative without improving gasterior core. Mossiuring overses of bibliod cultures may unintersticately lead to decreased excellebility of data recossesy to thick shafefor resistance over time; such difficult shafefor and the enrichment. | - | |
| Mathias and Baker, "Developing Qualit | and additional pair is inclumentation of common measures. ty Measures to Address Overuse, <i>JAMA</i> , May 8, 2013, p. 1897. | | |
| | | | |
| | | | |