

Quality: Measuring, Reporting & Rewarding Performance

Call for Papers

Public Reporting & Pay-for-Performance: Practical Challenges

*Chair: R. Adams Dudley, University of California,
San Francisco*

Monday, June 26 • 2:00 pm – 3:30 pm

● **Implications of Hospital Size for Uncertainty About 'True' Performance Rankings In a Hospital Pay-for-Performance Program: Lessons from the Premier Hospital Quality Project**

Gestur Davidson, Ph.D.

Presented By: Gestur Davidson, Ph.D., Senior Research Associate, Div. of Health Services Research & Policy, Univ. of Minnesota, 2221 University Ave. S.E. Suite 345, Minneapolis, MN 55414; Tel: 612-625-2339; Email: davidog64@umn.edu
Research Objective: We construct Bayesian statistical models estimated with data from Year I of the Premier Hospital Quality Incentive Demonstration (PHQID) project—sponsored by CMS—to demonstrate the impact that hospital size—by itself—is likely to have on our ability to infer “true” ranks from the data that would be generated and used in pay-for-performance programs patterned after the one being used in this demonstration.

Study Design: We compare the widths of the 95% credible intervals for mean rank within our Bayesian models by hospital size, as measured by the number of patients seen per year in each condition. We also examine estimated probabilities of a hospital being in the top 20% of ranked hospitals as a percentile-specific rank estimator that is used within the PHQID.

Population Studied: We focus on the Premier hospitals participating in the PHQID plus an augmented sample of very small Critical Access Hospitals (CAHs) that participated in the Hospital Compare program of CMS. For these hospitals we focus on their composite scores in three medical conditions of Heart Failure (HF), Acute Myocardial Infarct (AMI), and Community-Acquired Pneumonia (CAP), where these composite scores measure the overall percentage of patients needing the indicated services who are reported to have received them.

Principal Findings: For small hospitals with relatively few numbers of patients needing the indicated services in these conditions there would be 4-6 times more uncertainty concerning their ‘true’ ranks than would exist for the largest of the hospitals. This uncertainty varies by medical condition, as expected, due to differences in the heterogeneity of composite scores among hospitals across these conditions.

Conclusions: A key recommendation found in all the literature on the policy-uses of ranks is the absolute necessity of accompanying estimates of rank/percentile placement with adequate measures of the uncertainty of those estimates. The findings from our study underscore the critical importance of this.

Implications for Policy, Delivery, or Practice: Even with adequate accompanying measures of the uncertainty surrounding estimated ranks, identifying relative quality from simple ranks based on composite scores will have well-known limitations that will impact smaller institutions more severely than larger ones. Alternatives to simple ranks to identify relative quality should be explored and evaluated before implementing full pay-for-performance programs.

Primary Funding Source: HRSA

● **Financial Incentives and Gaming in Alcohol Treatment**
Mingshan Lu, Ph.D., Ching-to Albert Ma, Ph.D.

Presented By: Mingshan Lu, Ph.D., Associate Professor, Department of Economics, University of Calgary, 2500 University Dr NW, Calgary, T2N1N4; Tel: (403) 220-5488; Fax: (403) 282-5262; Email: mingshan01@yahoo.com

Research Objective: The concern with cost and quality in the health care market has led to many innovations. In a recent paper, Rosenthal et al. (2004) documented the recent trend of “paying for performance,” the use of financial incentives for quality improvements. It is thought that if decision makers can be held accountable, efficiency will improve.

Accountability requires the availability of information so that appropriate rewards and penalties can be applied. Information gathering on utilization and health outcomes has therefore become a critical component for health care accountability. Although health care plans have been collecting utilization, quality, and performance data for some time, the use of these data for rewards is quite recent. Once this information is used for structuring incentives, can health plans continue to rely on information reported by providers? Clinician gaming, the manipulation of at least some aspects of information to enhance practitioners' own agendas, is a critical issue. As clinicians become more aware of the consequences of accountability and the incentives involved, would gaming be more common? Are information manipulations influenced mainly by financial incentives, or other factors? What determines the extent of gaming? Our paper provides direct testing of the impacts of financial incentives under paying-for-performance on clinician gaming.

Study Design: We make use of two data sets on client information in alcohol treatment episodes. Each of the two data sets originated from the same clinician. One set consisted of reports required by the Maine Addiction Treatment System (MATS) for administrative and funding assessments; the other was abstract from actual medical records. Performance Based Contracting (PBC) was a policy implemented by the state of Maine around the middle of the period for our data sets. MATS data were used by PBC to assess program performance, which then determined future funding. There are discrepancies in the two data sets. We hypothesize that the information inconsistencies are due to clinicians misreporting information to MATS to game the system. The implementation of PBC resulted in a financial incentive for clinicians to misreport information to MATS. Good performance recorded in MATS can be due to over-reporting of a client's alcohol use at admission, or under-reporting at discharge. We identify over-reporting and under-reporting by comparing MATS data against medical record abstracts. We identify gaming incentives by the time dummy of the implementation of PBC. We are able to directly test

gaming because of our unique data sets. The comparison between the MATS and medical abstract data is a straightforward way to identify gaming. The identification of financial incentives on gaming was due to the implementation of PBC.

Population Studied: About 1,000 alcohol abuse treatment episodes in the state of Maine covering the period from October 1990 to June 1995.

Principal Findings: We find strong empirical evidence for our hypotheses. PBC was found to have a significant and positive effect on clinicians over-reporting client alcohol use at admission and under-reporting at discharge. Because performance evaluation is from assessing alcohol use reduction and abstinence, these misreporting help boost treatment performance where clients actually have not improved.

Conclusions: We find that the introduction of PBC has increased gaming. The pattern of gaming is consistent with clinicians exaggerating treatment performance to obtain higher financial rewards. Our results call for attention on provider reactions against incentive mechanisms: information manipulation should not be ruled out when such actions have financial consequences.

Implications for Policy, Delivery, or Practice: Our study suggests two policy implications. First, and perhaps more directly, auditing should be used more often when regulatory authorities must rely on information supplied by providers for financial and funding decisions. Appropriate auditing may deter gaming, and gives more reliability to the veracity of reports. Second, and perhaps less directly, establishing a gold standard should be considered whenever it is feasible. In our case, the analysis was possible precisely because we were able to compare the administrative reports against an appropriate standard, namely the medical record abstracts. Having an independent and reliable data source for testing the reliability of administrative data may seem obvious but appears to have received less emphasis. Data collection methods should consider obtaining the same information in more than one way.

Primary Funding Source: NIAAA, the Alberta Heritage Foundation for Medical Research, Institute of Health Economics

●Measuring Outpatient Coronary Artery Disease Quality of Care Using Electronic Health Records: Pitfalls and Targets for Improvement

Stephen Persell, M.D., M.P.H., Jennifer M. Wright, M.D., Jason A. Thompson, BA, Karen S. Kmetik, Ph.D., David W. Baker, M.D., M.P.H.

Presented By: Stephen Persell, M.D., M.P.H., Assistant Professor, Medicine, Northwestern University, 676 N. St. Clair St. Suite 200, Chicago, IL 60611; Tel: 312-695-1524; Fax: 312-695-0951; Email: spersell@nmff.org

Research Objective: Electronic Health Records (EHRs) can support quality improvement and the public reporting of quality data. The Centers for Medicare and Medicaid Services (CMS) is evaluating the feasibility of using EHR-based quality indicators for outpatient care through the Doctor's Office Quality-Information Technology (DOQ-IT) project. These indicators, developed and maintained by the American Medical Association-convened Physician Consortium for

Performance Improvement (AMA/Consortium), may be used in future pay-for-performance programs. We sought to evaluate how well these measures reflect actual quality for coronary artery disease (CAD).

Study Design: We performed a retrospective chart review. We compared results of single automated review with a two-step process of automated review supplemented by physician review for apparent quality failures. The outcome measures were seven quality indicators for outpatient CAD patients: 1) antiplatelet drug, 2) lipid lowering drug, 3) beta blocker following myocardial infarction, 4) blood pressure measurement, 5) lipid measurement, 6) low-density lipoprotein cholesterol (LDL-C) control, and 7) angiotensin converting enzyme inhibitor for patients with diabetes or impaired left ventricular systolic function. We calculated the performance rates of the DOQ-IT indicators as follows: number meeting criteria / (number meeting criteria + number not meeting criteria with no exclusion criteria). We repeated these calculations after reclassifying patients based on physician chart review. Fifteen percent of charts were reviewed by two physicians. Inter-rater reliability for determining if misclassification occurred was good ($\kappa = 0.62-0.85$ for individual indicators, 0.78 overall).

Population Studied: Patients with diagnosis codes for CAD seen at an urban internal medicine practice using a commercial EHR.

Principal Findings: We identified 1006 patients with CAD diagnosis codes and at least 2 office visits in 2004. By automated review, adherence to the seven quality measures ranged from 73.0% for lipid lowering drug prescribing to 97.6% for blood pressure measurement. However, review of physician notes showed that many of the cases that appeared to fail the quality measures were misclassified. The percentage of apparent quality failures that subsequently satisfied the measure based upon chart review (i.e., had the recommended intervention or met exclusion criteria) was 72, 67, 48, 79, 38, 15, and 33 percent, respectively. Success rates calculated using the two-stage process ranged from 87.1 for LDL-C control to 99.2 for blood pressure measurement. Reasons for misclassification included incorrect use of diagnosis codes, failure to record data meeting indicator criteria in searchable fields, and failure to capture permitted exclusions using automated searches.

Conclusions: In a setting where quality was generally high, apparent quality problems were frequently due to measurement error rather than poor care. Measuring outpatient CAD care using these indicators derived solely from an EHR may not produce accurate results in their current form.

Implications for Policy, Delivery, or Practice: Prevalent classification errors may pose an important obstacle to using these indicators for provider accountability. Improving how clinicians document chronic disease care, adding standardized codes to explain clinical exceptions, and modifying EHRs to improve chronic disease documentation may make EHR-based CAD quality indicators more accurate.

Primary Funding Source: AHRQ

●How Many Doctors Does It Take to Treat A Patient? The Challenges That Fragmented Care Poses for Pay-for-Performance

Hoangmai Pham, M.D., M.P.H., Deborah Schrag, M.D., M.P.H., Ann S. O'Malley, M.D., M.P.H., Peter B. Bach, M.D., MAPP

Presented By: Hoangmai Pham, M.D., M.P.H., Senior Health Researcher, , Center for Studying Health System Change, 600 Maryland Avenue, SW, Suite 550, Washington, DC 20024; Tel: 202-554-7571; Fax: 202-484-9258; Email: mpham@hschange.org

Research Objective: Pay-for-performance initiatives are based on the suppositions that patient care can be attributed to individual physicians through observing stable care patterns over time, and that most physicians will have a reasonable amount to gain through improved performance. We quantify the extent to which: 1) patient care can be attributed to individual physicians; 2) care relationships remain stable over time; and 3) physicians serve as the usual provider for a sizeable proportion of their patient panels.

Study Design: Analysis of claims patterns from 2000-2002 for 1.92 million Medicare fee-for-service beneficiaries treated by one of 7,216 physician respondents to the Community Tracking Study Physician Survey (2000-2001). We use well-accepted algorithms for attributing care to individual providers in beneficiary-level and physician-level analyses of Medicare claims to determine the stability of care relationships and the degree to which these relationships capture the totality of care delivered. We also use 2001-2002 Medicare Current Beneficiary Survey data linked to respondents' claims to compare beneficiary perceptions and claims-based determinations of the stability of usual care relationships. We excluded beneficiaries with disability or end-stage renal disease, "snowbirds," and those who entered Medicare managed care, nursing homes, or hospice during the study period. Our main outcome measures include the proportion of visits consumed by beneficiaries that are associated with their usual physicians and those billed for by physicians that are associated with their usual patients; and the degree of beneficiary switching of usual physicians over time.

Population Studied: Medicare fee-for-service beneficiaries who were patients of respondents to Round 3 of the nationally representative Community Tracking Study Physician Survey; and beneficiaries who responded to the 2001 or 2002 Medicare Current Beneficiary Survey.

Principal Findings: On average, beneficiaries annually saw 2.8 primary care physicians (PCPs) and 5.8 specialists, for a total of 8.7 physicians in 5.1 different practices. Care delivered by physicians other than usual physicians accounted for a mean of 41.5% of beneficiaries' evaluation and management (E&M) visits. Claims-based determination of switching of usual physicians disagreed with beneficiary perceptions of usual care relationships in 37% of cases. For 9% of the beneficiaries who reported stable relationships with their usual providers of a year or more, evidence of a continuing relationship with a usual physician was not apparent based on analysis of claims. Under claims-based attribution, 29.5% of beneficiaries switched usual physicians from year to year. In physician-level analyses, a mean of 59.7% of PCPs' patients were not their usual patients, representing 29.7% of their E&M visits. The analogous proportions were 89.8% and

67.2% for medical specialists. PCPs had a mean of 36.0% and medical specialists 65.2% of their usual patients switch to a different usual physician each year. Results were substantively similar when practices, rather than individual physicians, were examined.

Conclusions: Care relationships in fee-for-service Medicare are highly fragmented and unstable over time.

Implications for Policy, Delivery, or Practice: Pay-for-performance reliant on claims-based attribution thus fails to capture a significant proportion of care delivered, and faces obstacles to implementation. Alternative approaches may sidestep assignment of usual providers by proportionately ascribing responsibility to all of a beneficiary's physicians; or rely on prospective designations of usual care patient-physician relationships.

Primary Funding Source: NIA

●The Relationship Between Performance on Quality Indicators and Mortality Rates: Results from Medicare's Hospital Compare Report Card

Rachel Werner, M.D., Ph.D., Eric T. Bradlow, Ph.D.

Presented By: Rachel Werner, M.D., Ph.D., Assistant Professor, Medicine, University of Pennsylvania, 1230 Blockley Hall; 423 Guardian Drive, Philadelphia, PA 19104; Tel: 215-898-9278; Fax: 215-573-8778; Email: rwerner@mail.med.upenn.edu

Research Objective: In response to concerns about poor quality of care in US hospitals, the Centers for Medicare and Medicaid Services recently began reporting data on hospital performance on their website, Hospital Compare. It is unknown whether performance on these measures is related to hospital-level outcomes. Our objective was to examine the relationship between hospitals' performance on the quality indicators (QIs) used in Hospital Compare with risk-adjusted mortality rates at those hospitals.

Study Design: We used a Bayesian approach to compare hospital's performance based on Hospital Compare's quality indicators with hospital's average risk-adjusted mortality rates (RAMR). To do this, we applied Bayesian "shrinkage" to each hospital's "pass-rate" for each of the QIs, which weights the hospital's performance based on the number of patients, and thus the degree of uncertainty used to calculate that rate. For instance, hospitals that report their performance based on a smaller number of patients have greater uncertainty and hence their results are shrunken more towards the population average. Then, to test the relationship between each QI and disease-specific RAMR, we calculated hospital-level logistic regression models with each hospital's baseline performance modeled as a function of its RAMR.

Population Studied: We studied all hospitals nationally that participated in Hospital Compare for the first six months of 2004 (this data was published on Hospital Compare on April 1, 2005). Our data came from two sources. First, we used publicly available data on the 10 QIs initially included in Hospital Compare. These indicators cover three disease areas (acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia) and were used to determine the performance of each hospital. Second, for all hospitals included in Hospital Compare, we used MedPAR data to calculate average disease-specific inpatient and 30-day RAMR for each hospital.

Principal Findings: The number of hospitals included in the analysis ranged from 2,825 to 3,587 for each QI. Of the 10 QIs included in the analysis, there were 5 QIs related to care of patients with AMI, 2 QIs for CHF, and 3 QIs for pneumonia. Performance on all 5 AMI QIs were significantly associated with lower inpatient and 30-day AMI mortality rates. Among the 2 CHF QIs, 1 was significantly associated with lower inpatient and 30-day CHF mortality rates (ACE inhibitor for LVSD) and one was significantly associated with higher inpatient and 30-day CHF mortality rates (LV assessment). Among the 3 pneumonia QIs, 2 were significantly associated with lower inpatient pneumonia mortality rates (initial antibiotic timing and pneumovax), but were not associated with improved 30-day mortality. The third pneumonia QI (oxygenation assessment) was significantly associated with higher inpatient and 30-day pneumonia mortality rates.

Conclusions: While quality indicators for AMI are significantly associated with lower mortality rates, this relationship does not hold true for CHF or pneumonia. In fact, several of the QIs for CHF and pneumonia were associated with significantly higher mortality rates for those diseases.

Implications for Policy, Delivery, or Practice: Measuring and reporting performance has great potential for improving health care quality, but if measures are not correlated with lower mortality rates their role in quality improvement may be limited.

Primary Funding Source: No Funding

Call for Papers

Public Reporting & Pay-for-Performance: Experiences to Date

*Chair: Gary Young, Department of Veterans Affairs
& Boston University School of Public Health*

Tuesday, June 27 • 10:30 am – 12:00 pm

●Using Web-Based Quality Data to Choose a Primary Care Physician: Which Information Do Patients Rely on Most

Gary Fanjiang, M.D., MBA, Ted von Glahn, MSPH, Hong Chang, Ph.D., William H. Rogers, Ph.D., Dana Gelb Safran, ScD

Presented By: Gary Fanjiang, M.D., MBA, Fellow, Clinical Care Research, Tufts-New England Medical Center, 750 Washington Street, NEMC Box 63, Boston, MA 02114; Tel: (617) 636-5168; Fax: (617) 636-0525; Email: gfanjiang@hotmail.com

Research Objective: While there is growing demand to help patients make informed health care decisions by providing data on quality, studies suggest that available data is often ignored. This may be due, in part, to the types of quality data that have been available to date. Studies consistently find that patients are most interested in data to inform their physician choices, but scant physician-level quality data is available. This study examines whether patients seeking a new primary care physician (PCP) make use of web-based physician-specific information and evaluates which types of information users find most important.

Study Design: Patients were invited by mail to view web-based information about the physicians. Information included

physician credentials (years in practice, medical school, specialty certification, hospital affiliation), personal characteristics (age, gender, ethnicity, languages spoken), office location and hours, and patient-survey scores. Patient-survey scores were based on the Ambulatory Care Experiences Survey: Short-Form, a well-validated questionnaire that produces 5 summary measures: appointment access, interpersonal quality, coordination of care, health promotion and willingness to recommend the physician. After viewing the data (n=382), participants were asked to complete a questionnaire that asked them to indicate their preferred PCP and to evaluate the usefulness of the information presented. We examined the relative importance to patients of the different types of information, and evaluated the concordance between the information patients rated as most important and the characteristics of the physicians they preferred.

Population Studied: Adult patients seeking a new PCP at two California medical groups (n=2225).

Principal Findings: Seventeen percent of invited patients reviewed the web-based physician data. Patient-survey results were considered more important than each of the other types of information by approximately half of patients (range: 49% compared to M.D. personal characteristics, 61% compared to office convenience). Among the survey-based measures, those considered most important were willingness to recommend the physician (41%) and interpersonal quality (37%). Patients reported information priorities were highly concordant with the characteristics of their actual physician choice. For example, 84% of those citing “willingness to recommend” as most important selected the PCP with the highest or second highest score on this measure. Similarly, 88% of participants citing interpersonal quality as most important selected the PCP with the highest interpersonal quality score.

Conclusions: Approximately one in six patients seeking a new physician logged-on and used web-based data to inform their choice. Of the types of information available, patient survey results were widely regarded as a priority, with patients’ assessments of interpersonal quality and willingness to recommend valued most highly by prospective patients. Patients’ revealed preferences in choosing a physician were well aligned with their stated priorities.

Implications for Policy, Delivery, or Practice: With minimal outreach, web-based information concerning physicians attracted a sizable group of patients known to be seeking a new PCP. Survey-based information was valued highly by a majority of patients seeking to employ data to inform their physician choice. Moreover, users appear capable of using the information to accurately select a physician well-aligned with their stated priorities.

Primary Funding Source: RWJF

●**The Relationship between Pay-for-Performance Incentives and Quality Improvement: A Survey of Massachusetts Physician Group Leaders**

Ateev Mehrotra, M.D. M.P.H., Steven D. Pearson, M.D., MSc, Kathryn L. Coltin, M.P.H., Ken P. Kleinman, Ph.D., Janice Singer, M.P.H., MA, Eric C. Schneider, M.D., MSc

Presented By: Ateev Mehrotra, M.D. M.P.H., Post-doctoral Fellow, Dept of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115; Tel: 617-548-7517; Fax: ; Email: amehrotra@partners.org

Research Objective: Despite growing enthusiasm for pay-for-performance incentives, few studies have examined the scope of these incentives or the responses of physician groups. Our objective was to gather current information from leaders of physician groups about pay-for-performance incentives and to examine whether incentives are associated with greater use of quality improvement initiatives.

Study Design: Structured telephone survey of leaders of 79 of the 100 Massachusetts physician groups. We determined the prevalence of pay-for-performance incentives in physician group contracts with health plans and prevalence of physician group quality improvement initiatives. To assess the association between pay-for-performance incentives and quality improvement initiatives, we specified a single multivariable logistic regression model that encompassed the eight HEDIS® measures in our study. This model controlled for other factors that might influence the use of quality improvement initiatives.

Population Studied: All Massachusetts physician groups that include three or more primary care physicians.

Principal Findings: Most groups (89%) reported pay-for-performance incentives in at least one health plan contract. Incentives were most commonly tied to performance on HEDIS® quality measures (89% of all groups) and utilization measures (66% of all groups). Among groups with pay-for-performance the incentives accounted for 2.2% of overall revenue (range 0.25-8.8%) and 36% reported that pay-for-performance incentives were very important or moderately important to the group's financial success. Across the eight HEDIS® quality measures, we found an association between the presence of a pay-for-performance incentive on a measure and the use of a quality improvement initiative related to the measure (OR 1.6, p=0.04) after adjustment for other characteristics of physician groups.

Conclusions: Pay-for-performance incentives are now common among physician groups in Massachusetts. Although the scope and magnitude are still modest for most groups, we found an association between pay-for-performance incentives and use of quality improvement initiatives.

Implications for Policy, Delivery, or Practice: If this association is causal, these results support the idea that pay-for-performance incentives lead to an increased focus on quality improvement. Health plans are using pay-for-performance incentives for two purposes, to improve quality and to decrease costs.

Primary Funding Source: RWJF

●**Pay For Performance – The Impact on Patient Quality of Care in a Community Setting**

Irena Pesis-Katz, ABD, Peter Veazie, Ph.D., Norman Lindenmuth, M.D., Kathleen Curtin, Howard Beckman, M.D., Robert Greene, M.D.

Presented By: Irena Pesis-Katz, ABD, Medical Operations Research Project Manager, Medical Affairs, Excellus Blue Cross Blue Shield, 165 Court St., Rochester, NY 14647; Tel: (585)530-6741; Fax: (585)238-3658; Email: IrenaPesis-Katz@excellus.com

Research Objective: Pay for Performance (P4P) financial incentive programs are playing an ever-growing role in today's quality improvement efforts. The objective of this study is to evaluate the P4P impact on physicians' behavior in diabetes management.

Study Design: The study compares quality of care in Diabetes between three health plans using Health Plan Employer Data and Information Set (HEDIS) data. One plan, Excellus Blue Cross Blue Shield of Greater Rochester area (BCBSRA) served as the intervention group. The intervention consisted of a P4P program that incorporated (1) profiles with patient registries mailed to individual physicians three times a year, (2) interactive educational programs stressing measures of interest, and (3) year end distribution of dollars based in part on the diabetes measures. The control groups consisted of a competitor contracting with the same physicians in the same community (HMO A) and another Excellus HMO in a neighboring city (HMO B). A second analysis compares the rate of improvement in HEDIS scores between 2004 and 2003 for BCBSRA and each of the other two HMOs. HMOs A and B employed usual improvement methods in their environments. Study HEDIS measures included: yearly retinal eye exams, yearly HbA1c testing, poor HbA1c control (HbA1c>9.5), yearly LDL testing, LDL control (%LDL<100 or<130, and yearly monitoring of Nephropathy. The study uses HEDIS scores for the individual physician. Each physician's scores were compared to the specialty average and to a goal score. Individual performance profiles were sent to each physician starting October 2002. Financial incentives based on these profiles were allocated as of July 2003.

Population Studied: All diagnosed diabetes members defined by HEDIS inclusion criteria from the three commercial HMO plans in upstate NY.

Principal Findings: BCBSRA, with the P4P intervention, performed significantly better than HMO A or B. For 4 out of the 7 HEDIS diabetes measures BCBSRA had statistically significant increases in rates between 2003 and 2004. HbA1c testing increased from 88.3% to 91.7%, lipid profile increased from 90% to 95.1%, lipid control increased from 63.5% to 76.4%, and LDL control <100 increased from 38.9% to 48.9% (all are significant at the p<=0.05 level). HMO A and HMO B improved in only 1 out of 7 diabetes care measures. A comparison of the rate of improvement reveals that BCBSRA also performed significantly better in 6 out of 7 diabetes measures than HMO B (all but eye exams) and performed significantly better than HMO A in 4 out of 6 diabetes measures (eye exams, LDL profile, LDL control, and monitoring nephropathy).

Conclusions: This study suggests that financial incentives to physicians can contribute to improved quality of care, as measured by accepted HEDIS diabetes measures, when

compared to a non-financial control group. Furthermore, because HMO B and BCBSRA utilize the same physicians, the improved results in diabetes quality scores for BCBSRA suggest that physicians responded specifically to the P4P financial incentive and profile/registry reporting.

Implications for Policy, Delivery, or Practice: This study suggests that an individual physician's payment incentive programs can improve well established intermediate quality outcomes in patients with diabetes.

Primary Funding Source: No Funding

●Financial Incentives Work! Results of the CMS Hospital Quality Incentive Demonstration Project Year One

Denise Remus, Ph.D., RN

Presented By: Denise Remus, Ph.D., RN, VP, Clinical Informatics, Premier, Inc., 404 W Greenbriar, Dallas, TX 75208; Tel: 214-212-7774; Fax: 704-733-4821; Email: Denise_Remus@premierinc.com

Research Objective: The CMS/Premier Hospital Quality Incentive Demonstration (HQID) project was set up to examine the following hypothesis: Can economic incentives effectively improve quality of care?

Study Design: The study design is observational using both prospective and retrospective components; participants were self-selected and all had to be enrolled in the Premier Perspective™ database.

Population Studied: All adult patients (age > 18) admitted to participating hospitals within the four clinical conditions of acute myocardial infarction (AMI), heart failure (HF), pneumonia (PN) and coronary artery bypass graft (CABG) procedures; and all Medicare patients (primary or secondary payer = Medicare) admitted for hip or knee (HK) replacement procedures.

Principal Findings: The HQID uses data from individual quality measures, within each clinical area, to create an aggregate score or the Composite Quality Score (CQS) representing overall quality. The project incorporates both process (e.g., ASA on admission, oxygen assessment) and outcome measures (e.g., mortality, readmissions), 33 total (year 1) across the five areas. Analyses of final (validated) data from year 1 identified statistically significant improvements (all $p < 0.001$) from the first quarter of the project (Q4-03) to the fourth quarter of the project (Q3-04) in the CQS in all five clinical areas. Hospitals were placed in deciles based on CQS and the top 10% of hospitals, in each clinical area, received a 2% quality incentive payment from Medicare; the next 10% received a 1% bonus; \$8.85 million was awarded to 123 top performers in year 1. The top 50% received public recognition. HQID participant data was compared to data of non-participating hospitals from Hospital Compare for the time period of Q2-04 to Q1-05 for the 17 process measures similar across the two projects. Preliminary analyses found that the average measure rate (percent of patients receiving intervention) was significantly higher ($p < 0.001$) for HQID participants than non-participants in 10 measures, sig. higher ($p < 0.05$) in 2 measures, higher (ns) in 3 measures and lower (ns) in 2 measures; and sig. higher ($p < 0.001$) in the overall composite process score (82.2% vs 73.2%, HQID $n=259$, Hospital Compare $n = 3303$, limited to measure denominators > 24 cases). Preliminary data from year 2 demonstrates continued improvement. Additional results will be presented

including analysis of CQS by hospital characteristics and further evaluation of year 2 data.

Conclusions: The HQID project, incorporating both quality incentive payments and public recognition, resulted in statistically significant improvements in the first year of the project and, for the majority of comparative measures, average rates significantly higher than a national sample of publicly reported data.

Implications for Policy, Delivery, or Practice: The demonstration project has two more years of data to report, however initial findings based on performance in year 1 indicate that financial incentives, along with public recognition, resulted in statistically significant improvement in quality as measured by CQS in five clinical areas. While additional research is necessary to examine other factors associated with the overall improvement, the preliminary data supports efforts to link financial payments to quality.

Primary Funding Source: CMS, Premier, Inc.

●A Randomized Controlled Trial of an Educational and Motivational Intervention to Enhance Consumers' Use of Health Plan and Medical Group Quality Data

Patrick Romano, M.D., M.P.H., Julie A. Rainwater, Ph.D., Jorge A. Garcia, M.D., MS, Geeta Mahendra, Daniel J. Tancredi, MS, Janet M. Keyzer, RNC, MPA

Presented By: Patrick Romano, M.D., M.P.H., Professor, Division of General Medicine, University of California Davis, 4150 V Street; PSSB Suite 2400, Sacramento, CA 95817; Tel: 916.734.7005; Fax: 916.734.2732; Email: psromano@ucdavis.edu

Research Objective: Health care quality reports are increasingly prevalent but have little impact on consumer behaviour; many consumers appear not to understand quality data or fail to appreciate its salience.

Study Design: A cluster-randomized, controlled trial of a two-pronged educational/motivational intervention to enhance use of quality data during Open Enrollment: (1) a mailing with the California HMO Report Card, California's HMO Guide, and a motivational letter "negatively framed" to arouse concerns about health care quality; and (2) toll-free telephone and e-mail hotlines staffed by counselors trained to provide advice around enrolment decisions. Both components were designed to motivate and empower consumers, based on guidance from focus groups and previous research. Individuals in the "usual care" group received no mailings, but had access to standard print and Internet resources.

Population Studied: PacAdvantage is a non-profit purchasing pool offering 10 managed care plans to small businesses (2 to 100 employees) in California. We generated a stratified random sample of brokers offering PacAdvantage to small business clients during May-July 2003 (oversampling small brokers and brokers likely to have higher switch rates among eligible employees). Brokers were blindly allocated to either the control or intervention group; all of their clients were allocated similarly. Eligible employees (EEs) in both groups were surveyed by mail within 2 months, and their choices were captured from enrolment data. We compared plan-switching between the intervention and control groups, overall and after categorizing switches as quality-increasing, quality-decreasing, or neutral. We also compared self-reported use of quality information, reasons for switching, outcome expectations,

contemplation of switching, ease of selecting a plan, and self-efficacy. All analyses were weighted and adjusted to account for the sampling design.

Principal Findings: 292 brokers with 1,835 EEs were randomized to the intervention group; 246 brokers with 1,578 EEs were randomized to the control group. About 30% of EEs in the intervention group, and 37% of EEs in the control group, dropped out of PacAdvantage. Only 22 intervention group members used the advice line and 3 used the e-mailbox, with a broad array of questions and concerns. By the end of Open Enrollment, 9.2% of intervention and 7.0% of control group members switched plans (NS). 21% of intervention group switchers versus 35% of control group switchers moved to a higher rated plan; 27-28% in both groups moved to a lower rated plan. Intervention group members were more likely to have considered switching (35% versus 28%) and to have reviewed information about health care quality (38% versus 8%) than control group members. However, intervention group members were more likely to report “a big problem” finding a suitable plan than control group members (15% versus 9%), and switchers in the intervention group were more likely to express concern that their quality of care would suffer (8% versus 1%).

Conclusions: Educational/motivational interventions designed to increase perceived benefits and decrease perceived barriers, with negative framing, may increase consumers’ use of quality information but are unlikely to affect actual choices.

Implications for Policy, Delivery, or Practice: Consumers are reluctant to act on well-presented quality information, given competing concerns about cost and access, and such information may even trigger some distress.

Primary Funding Source: AHRQ

Related Posters

Quality: Measuring, Reporting & Rewarding Performance

Poster Session B

Monday, June 26 • 5:30 pm – 7:00 pm

•Variability in Orthopaedic Practice Patterns in the United States

Kevin Bozic, M.D., M.B.A., Amanda Smith, M.P.H., Harold Luft, Ph.D.

Presented By: Kevin Bozic, M.D., M.B.A., Associate Professor, Orthopaedic Surgery, UCSF, 500 Parnassus Ave., MU-320W, San Francisco, CA 94143; Tel: 415-476-3900; Email: bozick@orthosurg.ucsf.edu

Research Objective: Recent data suggests that adherence to evidence based practice guidelines reduces costs and improves quality in health care. Very few evidence based practice guidelines exist in orthopaedic surgery. The purpose of this study was to review the variability in practice patterns in the United States for commonly performed orthopaedic surgical procedures to determine if adherence to established practice guidelines is associated with improved patient outcomes (e.g., fewer operative complications) and lower overall cost of care.

Study Design: Administrative claims data were collected for all orthopaedic procedures completed by UnitedHealth Group providers (n=48,167) between January 1, 2002 and December 31, 2003. Nine common orthopaedic surgical procedures were selected for review for adherence to five generally accepted clinical practice guidelines, including the peri-operative utilization of imaging studies and physical therapy. Differences in adherence between geographic regions were compared using summary measures normalized for severity of illness. Total cost of care included all costs incurred from the orthopaedic surgical procedure (including inpatient costs) and costs from all associated orthopaedic procedures within one year peri- and post-operative.

Population Studied: Claims data for 116,810 patients in 26 discrete geographic regions of the United States were reviewed.

Principal Findings: Adherence to practice guidelines and total cost of care varied significantly by both procedure type and geographic region for all nine orthopaedic surgical procedures. Increased adherence to generally accepted clinical practice guidelines was not always associated with lower total cost or improved clinical outcomes.

Conclusions: Significant variation exists in adherence to generally accepted orthopaedic clinical practice guidelines by procedure type and geographic region of the country. Further study is necessary to define evidence based practice guidelines, and to objectively investigate the correlation between adherence to evidence based practice guidelines, clinical outcomes, and cost of care.

Implications for Policy, Delivery, or Practice: Very few evidenced-based practice guidelines currently exist in the field of orthopaedic surgery. As a result, orthopaedic practice patterns vary considerably, leading to marked differences in clinical outcomes and cost of treatment in this resource-intensive specialty. The results of our study suggest that adherence to generally accepted clinical practice guidelines is not necessarily associated with improved quality and lower costs. These findings underscore the need to develop evidence-based treatment algorithms based on Level I evidence.

Primary Funding Source: No Funding

•Hospital Quality: Correlation Among Process-of-Care Measures and Relationship with Mortality

Elizabeth Bradley, Ph.D., Jeph Herrin, Ph.D., Brian Elbel, M.P.H., Robert L. McNamara, M.D., MHS, John A. Spertus, M.D., M.P.H., Harlan M. Krumholz, M.D., SM

Presented By: Elizabeth Bradley, Ph.D., Associate Professor, Epidemiology and Public Health, Yale School of Medicine, 60 College Street, New Haven, CT 06052; Tel: 203.785.2937; Fax: 203.785.6287; Email: elizabeth.bradley@yale.edu

Research Objective: Hospitals vary substantially in the quality of care they provide for patients with acute myocardial infarction (AMI), and many patients do not receive guideline-based care. As part of the national effort to improve hospital quality, the Centers for Medicare & Medicaid Services (CMS) monitor and publicly report hospital performance on 7 core AMI process measures. Despite substantial investment in this national quality monitoring system, little is known about how different core process measures track with each other or about how strongly hospital performance on the core process

measures is associated with hospital short-term mortality rates for patients with AMI. Using a hospital-level analysis, we sought to determine the correlation among the 7 core process measures and then to evaluate how much of the variation in hospitals' 30-day risk-adjusted mortality rates was explained by hospital performance on the core AMI process measures.

Study Design: We estimated hospital performance on each of the 7 AMI core process measures using National Registry of Myocardial Infarction (NRMI) data from hospitals reporting AMI admissions in 2001 and hierarchical generalized linear modeling (HGLM). We calculated hospital-specific, risk-adjusted 30-day mortality rates with data from CMS Medicare claims using a risk-adjustment model recently endorsed by the National Quality Forum. We used standard correlation and linear regression analyses to examine the relationship between process and outcome performance at the hospital level.

Population Studied: The sample for the main analysis included 96,669 patients with acute myocardial infarction aged 66 years or older and treated in 1,022 of the 1,112 NRMI hospitals that could be matched to hospitals in the CMS Medicare claims database.

Principal Findings: We found strong correlations (correlation coefficients = 0.41-0.75; P-values < 0.001) for all pair-wise comparisons between beta-blocker use at admission and discharge, aspirin use at admission and discharge, and angiotensin converting enzyme inhibitor use at discharge. Weaker, but still statistically significant, correlations between these admission and discharge medication performance measures and the time to reperfusion therapy and smoking cessation counseling measures (correlation coefficients = 0.31 – 0.40; P-values < 0.001). Hospital performance on core AMI process measures was weakly correlated with hospital performance in risk-adjusted 30-day mortality rates, explaining only 6.5% of the hospital-level variation in risk-adjusted mortality rates for patients with AMI.

Conclusions: Of the 7 core process measures for AMI, medication-based measures are strongly correlated with each other, but less correlated with system-related processes like timely reperfusion therapy. Together, these measures explained relatively little of the hospital-level variation in short-term, risk-adjusted mortality rates for patients with AMI.

Implications for Policy, Delivery, or Practice: Although the HQA processes-of-care are important in pursuing improved AMI quality, they are less useful for making inferences about a hospital's risk-adjusted, short-term mortality rate. Until we develop additional process measures that explain more of the variation, reporting not only the current core measures but also short-term mortality rates, after accounting for case mix and random variation, is a reasonable approach to characterize the hospitals' overall quality of care.

Primary Funding Source: National Heart, Lung, and Blood Institute; Donaghue Medical Research Foundation

●Health Care Satisfaction for Medicaid Managed Care Enrollees: The Role of Beneficiary Experiences

Joseph Burton, M.S., John Kautter, Ph.D.

Presented By: Joseph Burton, M.S., Senior Researcher, RTI International, 411 Waverley Oaks Road, Waltham, MA 02452-8414; Tel: 781-434-1710; Fax: 781-434-1711; Email: joeburton@rti.org

Research Objective: This study investigates how individual characteristics and personal experiences with health plans and providers affect the ratings of health plans, doctors, and overall health care in a Medicaid Managed Care population. Previous research has shown that individual characteristics, including health status, affect measures that are part of the Consumer Assessment of Health Plans Survey (CAHPS®). These analyses use CAHPS® items to identify how personal experiences influence plan and provider ratings independent of individual characteristics.

Study Design: Satisfaction measures based on CAHPS® are designed for and reported at the health plan level. We analyzed individual-level responses to the 2000 Ohio Medicaid Managed Care Consumer Satisfaction survey, which included items from CAHPS®. We developed person-level experience scales based on the standard CAHPS® composite measures and modeled their effects on the ratings of health plans, providers, and overall health care. We grouped the ratings into “High”, “Medium” and “Low” and, using ordered logit models, regressed the ratings on individual characteristics and the personal experience scales. We restricted the analysis to beneficiaries who had personal doctors and who visited a doctor's office or clinic in the prior six months. We also conducted separate analyses for adults and children.

Population Studied: Medicaid managed care beneficiaries.

Principal Findings: Three of the four scales—experiences with getting needed care (Access), communicating with doctors (Communication), and courteousness of office staff (Courtesy)—were highly significant in explaining plan and provider ratings. Adults who reported problems getting needed medical care had less than half the odds of rating their health plan favorably. Respondents who viewed doctors as good communicators had 50 percent higher odds (in the case of children) and 80 percent higher odds (in the case of adults) of rating health plans favorably. The perceived communication skills of providers had a larger effect on the rating of personal doctors: respondents who thought their doctors communicated well had twice the odds of rating them favorably and over three times the odds of rating their children's doctors favorably. The rating of personal doctors also influenced—more than any other factor—how beneficiaries rated health plans and overall health care. When parents rated their children's doctors “High” compared to “Low”, they had five times higher odds of rating health plans favorably and 26 times higher odds of rating their children's overall health care favorably.

Conclusions: This study revealed that personal experiences directly influence how Medicaid managed care beneficiaries or their parents rate health plans, personal doctors, and overall health care, independent of health status and demographic characteristics. In addition, how beneficiaries perceive their doctors is the most important determinant of plan satisfaction.

Implications for Policy, Delivery, or Practice: These findings provide empirical support for the current trend among researchers to develop physician-based performance and satisfaction measures. Doctors and staff at doctors' offices are the focal point of receiving primary care, and satisfaction measures among Medicaid beneficiaries in Ohio appear to be sensitive to perceptions and experiences with doctors and their staffs.

Primary Funding Source: CMS

•Pay for Performance or Penalize for Financial Sustainability? -- An Implication of the Relationship between Hospital Quality and Financial Performance

Li-Wu Chen, Ph.D., Jeff Sun, Ph.D., Wanqing Zhang, M.Ed., Keith Mueller, Ph.D., Roslyn Fraser, M.A.

Presented By: Li-Wu Chen, Ph.D., Associate Professor, Preventive & Societal Medicine, University of Nebraska Medical Center, 984350 Nebraska Medical Center, Omaha, NE 68198-4350; Tel: (402)559-5260; Fax: (402)559-7259; Email: liwuchen@unmc.edu

Research Objective: Little has been written about the relationship between hospital financial and quality performance. A better understanding of this relationship would provide useful insight into the effectiveness of rewarding (or penalizing) hospitals using pay-for-performance mechanism. This study intends to address this issue with respect to the quality of hospital inpatient care for cardiovascular disease patients.

Study Design: We use a logistic regression model to examine the effect of prior hospital financial performance on patient's later likelihood of death during hospitalization for patients who have a principal diagnosis of Acute Myocardial Infarction (AMI) or Congestive Heart Failure (CHF). A Generalized Estimating Equations (GEE) approach is used to account for the correlation among patients from the same hospital. The use of prior hospital financial performance measures to predict patient's later mortality would help alleviate the potential problem of having a reverse causal relationship between quality and finance. Our model controls for patient-level (e.g., demographic, APR-DRG severity and comorbidity measures) as well as hospital-level (e.g., ownership, capacity, staffing, cardiovascular treatment resources) characteristics.

Population Studied: This study uses a national sample of hospitalized patients with AMI or CHF. Patients' hospital discharge data from the 2002 Nationwide Inpatient Sample (NIS) of the Healthcare Cost and Utilization Project (HCUP) are merged with hospital-level data from the American Hospital Association annual survey and the Medicare Cost Report. Separate models are estimated for AMI patients and CHF patients, as well as for rural hospitals and urban hospitals.

Principal Findings: Our results suggest that prior hospital financial performance affects in-hospital mortality for AMI and CHF patients in rural hospitals, but not in urban hospitals. Among rural hospitals, a greater prior hospital return-on-equity is associated with a lower likelihood of in-hospital death for both CHF and AMI patients. In particular, CHF and AMI patients treated by rural hospitals with a higher prior return-on-equity by 10 percentage points are 5% ($p < 0.01$) and 6% ($p < 0.1$) less likely to die within hospitals, respectively, as compared to their counterparts. In addition, AMI patients who

are treated by rural hospitals, which have a higher prior equity financing ratio, are less likely to die in hospitals ($p < 0.05$) than their counterparts. On the other hand, prior hospital margin and current ratio are found to be positively associated with the likelihood of in-hospital death ($p < 0.05$) for CHF patients in rural hospitals (but insignificant results for AMI patients).

Conclusions: Quality performance of treating AMI and CHF patients is more sensitive to prior hospital financial performance for rural hospitals than for urban hospitals. Our findings suggest that rural hospitals, given limited resources, are more likely to perform better in quality (e.g., through investment in equipment and personnel) if their financial stability and capability of making good returns out of investment (i.e., equity) is better. Otherwise, rural hospitals without considerable financial stability may need to sustain their operation through enhancing profitability and short-term liquidity at the expense of quality performance.

Implications for Policy, Delivery, or Practice: Policy makers should be careful in designing pay-for-performance mechanisms, especially in the case of rural hospitals. Our research suggests that policy makers should take rural hospitals' financial conditions and potentials into consideration when designing the quality criteria for rewarding or penalizing hospitals. A pay-for-performance system without such consideration may jeopardize the financial viability of some rural hospitals, thus affecting the access to hospital care for many rural Americans.

Primary Funding Source: HRSA

•Developing a Spanish-language Sample Report for CAHPS®

Kathryn Pitkin Derose, Ph.D., M.P.H., David E. Kanouse, Ph.D., Beverly Weidmer, MA, Robert Weech-Maldonado, Ph.D., M.B.A., Rosa-Elena Garcia, M.P.H., Ron D. Hays, Ph.D.

Presented By: Kathryn Pitkin Derose, Ph.D., M.P.H., Social Scientist, Health Program, RAND Corporation, 1776 Main St PO Box 2138, Santa Monica, CA 90407-2138; Tel: (310) 393-0411 ext 6302; Fax: (310) 260-8161; Email: pitkin@rand.org

Research Objective: Develop a culturally appropriate sample report on health plan quality for Spanish-speaking Latinos for the Consumer Assessments of Healthcare Providers and Systems (CAHPS®) project.

Study Design: We followed a procedure of multiple forward translations, review by committee, and readability or literacy assessment to produce a draft Spanish report for usability testing. We conducted cognitive interviews with 16 Spanish-speaking Latinos and used the results to revise the report for additional testing with 8 more Spanish-speaking Latinos. Participants represented a wide range of educational backgrounds, but all self-reported feeling comfortable with reading written materials.

Population Studied: Twenty-four Spanish-speaking Latino men and women from Los Angeles, California (8 Mexican American, 3 El Salvadoran, 1 Nicaraguan) and Ocala, Florida (8 Puerto Rican, 4 Cuban).

Principal Findings: Of the 24 participants, the median age was 41 years (range=19-59 years), 67% were female, 62% had less than a high school education (21% < 6th grade), and the median number of years in the U.S. was 8.5 (range=0.5-32 years). Overall, respondents understood and would use such a report to choose a health plan. However, nearly all stressed

that health care costs, access, and coverage information was essential in evaluating the quality of a health plan. Respondents with less formal education had more difficulty understanding the graphical information than those with more formal education--in particular segmented bar graphs that presented data on the proportion of health plan members who responded in each of three response categories. Changing the bar graphs to simple horizontal bar graphs with dichotomous splits improved overall comprehension by respondents. The addition of text under each graph summarizing the overall message of each graph did not improve comprehension. A summary chart that compared all health plans across all dimensions proved easier to comprehend than individual bar graphs, especially when the stars were changed to word icons. Almost all respondents, regardless of education, had trouble understanding comparisons of individual health plans to the average of all health plans.

Conclusions: This study of Spanish-speakers indicates a similar level of understanding of reports about health plan quality as found in prior studies of English-language populations. Concepts and terms referring to aspects of quality appear to translate well between English and Spanish, suggesting equivalence in translatability and comparability of the information provided. The results also suggest that health plan quality information needs to be integrated with information about costs, access and coverage to facilitate decision-making. Simplifying graphical information involves losing some detail but makes information more usable. Summary charts facilitate across plan comparisons, but relative differences are hard for consumers to understand.

Implications for Policy, Delivery, or Practice: Difficulties in comprehension of quality information observed for Spanish-speaking participants are similar to those previously observed in cognitive and usability testing for English-speakers. As such, they appear to derive from difficulties inherent in the subject matter along with participants' limited experience with some of the concepts and presentational strategies rather than factors specific to Spanish language or Latino culture.

Primary Funding Source: AHRQ

●A Comparative Analysis of Chronic and Non-Chronic Insured Commercial Member Cost Trends

Ian Duncan, FSA, MAAA, Iver Juster, M.D., Robert Bachler, FCAS, FSA, MAAA, Rebecca Owen, FSA, FCA, MAAA

Presented By: Ian Duncan, FSA, MAAA, President, Solucia Inc., 1477 Park Street, Suite 316, Hartford, CT 06106; Tel: 860-951-4200; Fax: 860-951-4204; Email: iduncan@soluciaconsulting.com

Research Objective: To compare the medical cost trends of members with chronic conditions, non-chronic members and all members of commercial health plans over time.

Study Design: Chronic and Non-chronic cost trends were calculated annually for 1999 through 2002 and were compared under different assumptions regarding identification of chronic disease and medical services.

Qualification for the Chronic group was defined as having coronary artery disease, heart failure, diabetes, asthma or chronic obstructive lung disease.

Population Studied: A commercially-available data set of 1.5 million commercially insured members. The dataset was purchased from Ingenix, Inc. of Minneapolis, MN.

Principal Findings: When Chronic and Non-chronic members are identified and included in the population prospectively, the average 3-year trend over the study period for chronic and non-chronic members adjusted for high cost outliers were 4.9% and 13.9% respectively. Adjusting the groups for differences in service mix had little impact on the divergence in trends. Similarly, excluding extreme high-cost outliers had little impact on the difference in trends. However, altering the Chronic selection algorithm to eliminate migration between groups (thus classifying a member as always Chronic if identified as Chronic at any point in the four years) caused the trends to converge.

Conclusions: Estimating Disease Management program financial outcomes based on the assumption that, absent the program, the Chronic population would have had the same trend as the Non-chronic population can lead to erroneous conclusions. Trends calculated using a prospective identification methodology introduce an (upward) bias into estimates of outcomes. We refer to this effect, which has not previously been described or discussed in the literature, as "migration bias".

Implications for Policy, Delivery, or Practice: The trend-adjusted historical control methodology is the most prevalent methodology used in Disease Management outcomes. Our research indicates that this method is at risk of significant upward bias in the estimate of financial outcomes because of the way in which the trend adjustment is calculated, which in turn calls into question many of the published disease management population savings estimates.

Primary Funding Source: Society of Actuaries

●Measuring Staff Perceptions on Patient Safety in Order to Focus Safety Improvement Resources on areas with the Greatest Need

Paula Edwards, BS, Patricia Richardson, Tracy Scott, Ph.D., James Jose, M.D., Kimberly Rask, M.D., Ph.D., François Sainfort, Ph.D.

Presented By: Paula Edwards, BS, Researcher, Health Systems Institute, Georgia Institute of Technology, 1258 Dekalb Ave #134, Atlanta, GA 30307; Tel: 678-613-7650; Email: pjedwards@gatech.edu

Research Objective: To measure patient safety culture among patient care staff at a two-campus pediatric hospital system serving a large metropolitan area in order to identify patient safety areas in need of improvement. The identified areas of need were then used to prioritize patient safety initiatives across the system to concentrate resources and efforts on areas with the greatest need.

Study Design: The Westat Hospital Survey on Patient Safety was administered to patient care staff at the hospital system. The Westat survey uses 42 Likert-scale questions to collect perceptions regarding 12 dimensions of safety culture. Paper versions of the survey were supplied along with pre-printed envelopes for respondents to return surveys via internal mail. All responses were anonymous.

Population Studied: The patient safety survey was distributed to 1530 patient care staff at two pediatric hospitals (one teaching hospital and one community hospital). 369 survey responses were received (24.1%). These responses included registered nurses (68.8%), respiratory therapists (14.1%), patient care technicians (7.6%), and other staff (9.5%).

Responses were approximately evenly distributed between the two campuses.

Principal Findings: Scores on each of the 12 safety dimensions were constructed from survey responses for each respondent. Scale scores range from 1 (very poor) to 5 (excellent). Mean scores on the scales ranged from 3.08 to 3.90, indicating generally favorable perceptions of patient safety among the staff surveyed. An ANOVA test with Tukey post-hoc tests compared scores across dimensions to identify strengths and weaknesses across the system ($F=31.991$, $p<0.001$). Strengths identified were Teamwork within Hospital Units, Organizational Learning, Hospital Management Support for Safety, and Supervisor Expectations/Actions. Scores on each of these dimensions were significantly higher ($p=0.02$) than seven or more other dimensions. Areas for improvement were identified as Nonpunitive Response to Error, which had significantly lower scores than all other dimensions ($p<0.01$), as well as Hospital Handoffs & Transitions and Teamwork Across Hospital Units, each of which was significantly lower than several other dimensions ($p=0.01$). Similar analyses of each campus confirmed that these strengths and weaknesses were consistent across both campuses and identified Communication about Error as an additional area for improvement at the community hospital. A t-test comparison of scores between campuses revealed that the teaching hospital scored slightly higher on three dimensions ($p<0.05$, difference in means ranged from 0.14 to 0.21 points): Feedback & Communication About Error, Communication Openness, and Organizational Learning. Based on these results, four initiatives are currently underway to improve patient safety in several of the identified areas for improvement. These initiatives are: initiating safety rounds, regular safety updates at existing staff meetings, a transfer of care check sheet and implementation of SBAR as a system communication tool.

Conclusions: The Hospital Survey on Patient Safety provided an effective measure of staff perceptions regarding patient safety at a pediatric hospital system. The survey results enabled this hospital system to identify areas for improving patient safety across the system as well as identifying individual differences between the two campuses.

Implications for Policy, Delivery, or Practice: The insight gained by this survey can help hospitals concentrate patient safety improvement efforts on areas in need of the most improvement.

Primary Funding Source: AHRQ

•Hospital Quality: What's Pricing & Markets Got to Do with It?

Ronald Fisher, Ph.D. (candidate), Ramesh Shukla, Ph.D., Robert Hurley, Ph.D.

Presented By: Ronald Fisher, Ph.D. (candidate), Research associate, Health Administration, Virginia Commonwealth University, P. O. Box 980203, Richmond, VA 23298-0203; Tel: 804-873-7844; Fax: 804-828-1894; Email: rlfisher@vcu.edu

Research Objective: Profound change towards more market-oriented governance has transformed US healthcare in the last decades. Though research has characterized many attributes of this social movement, uncertainty persists as to whether objectives such as more efficient pricing and market differentiation are, in fact, being realized. The study aims to

contribute to empirical findings that relate hospital quality performance to economic market factors. Three research questions were addressed: 1) Is economic competition raising the bar for quality performance? 2) Does quality cost more? And, 3) Is hospital quality performance differentiated as rationally expected by supply and demand relationships.

Study Design: A four year longitudinal multi-level design was employed to model relationships between hospital quality and economic performance. A quality performance measurement model was developed through selection of indicators that met statistical and face validity criteria. Four risk-adjusted indicators were found acceptable and included three AHRQ patient safety measures (PSIs) and a rather global, if qualified, mortality rate. Factor analysis justified the use of two indexed measures (e.g., mortality and errors) comprising two indicators each. Economic measures included both revenue and production function expense models to assess pricing mechanism effects on quality outputs. Local market structure factors of competition, health plan purchasing leverage, and market wage index were also modeled within the hierarchical dataset.

Population Studied: Acute hospitals operating in Virginia during 1998 and 2001.

Principal Findings: Significant between hospital variance was demonstrated, with hospital specific effects accounting for approximately 50% of the total variance of both quality performance measures. No significant trend of quality improvement was found for either quality measure, though the trajectory indicated a general lowering of Error events over time. Comparisons between economic and market dynamic models and "empty models" did not reveal any expected market effect on the variance component of the growth curve. No pricing mechanism fixed-effects were evidenced. Hospital quality does not cost more, even when market structure dynamics are accounted for. A significant relationship between "competition" and Error performance was found. Distribution of market share amongst a number of hospitals in a local market was associated with lower Error rates. Some hospital characteristics included as economic control variables, such as staffed bed occupancy rates, were found to positively associate with the adverse events; though the evidenced associations are not readily explained by expected market efficiency assumptions.

Conclusions: Validation is provided for the measurement of hospital quality outcome indicators as revealing significant hospital specific effects. The rationally expected utility of increasing reliance on market-oriented governance, at least as assessed by the pricing and market differentiation of quality outputs, has yet to be revealed. The finding continues to support the notion of healthcare quality performance as a "public good," undifferentiated by expected pricing mechanisms.

Implications for Policy, Delivery, or Practice: While neoclassic theory has shown to be useful tool in many economic situations and industrial sectors, healthcare research and policy may find guidance from alternative theoretical and methodological approaches of merit. The empirical findings of this study, considered as stylized facts, suggest Herbert Simon's adaptive rationality or institutional economic theory may be more productive.

Primary Funding Source: No Funding

●Reliability of Claims and Encounter Data to Capture Colorectal Cancer (CRC) Screening

Min Gayles Kim, M.P.H., Russell Mardon, Ph.D., Philip Renner, M.B.A.

Presented By: Min Gayles Kim, M.P.H., Senior Health Care Analyst, Quality Measurement, National Committee for Quality Assurance, 2000 L ST NW, Suite 500, Washington, DC 20036; Tel: 202-955-1731; Fax: 202-955-2599; Email: gayles@ncqa.org

Research Objective: To evaluate the reliability of claims/encounter data to identify colorectal cancer screening in a managed care population. It was hypothesized that due to varying periodicities of screening methods recommended in clinical guidelines, many of which are longer than average health plan enrollment tenure, and the limitations of claims/encounter data, a performance measure of CRC screening would require medical record review to accurately capture all screening modalities.

Study Design: A CRC screening measure was implemented by the National Committee for Quality Assurance (NCQA) as part of the Health Plan Employer Data and Information Set (HEDIS) in 2004. The measure assessed whether adults 50-80 years of age have had appropriate screening for colorectal cancer, consistent with national guidelines.

Population Studied: Plan level claims/encounter and medical record data were collected for commercial and Medicare plans of enrollees 50-80 years of age. Of 300 commercial and 161 Medicare plans that participated, 90 percent submitted valid rates for this measure.

Principal Findings: Denominator prevalence was high at 500 per 1000 enrolled members for commercial plans and nearly 800 for Medicare plans. The mean commercial rate was 41 percent for the plans that submitted data from claims/encounter data and 51 percent for plans that submitted data from claims/encounter plus medical record review with an overall mean of 47 percent. Medicare rates were about two percentage points higher. Of those that met the screening criteria, two thirds had evidence of colonoscopy, one third had received FOBT, one fourth had evidence of flexible sigmoidoscopy, and five percent had received DCBE. Screening rates were highest in the Northeast region (60%) of the U.S. and lowest in the South Central region (42%). In plans that collected both claims/encounter and medical records, about 80% of numerator hits were captured by claims/encounter data in both commercial and Medicare plans. This rate of administrative data capture was consistent across low to high performing health plans.

Conclusions: Results indicate a high proportion of CRC screening is captured by claims/encounter data. Health plans can use administrative claims/encounter data to support reporting and quality improvement of colorectal cancer screening. Over 80% of the data contributed came from claims/encounter data. This proportion was consistent across levels of performance on the measure, implying that better capture of claims/encounter data is not the source of variation in performance. Of the screening modalities, colonoscopy is the most frequently utilized method of screening. Despite compelling evidence that routine CRC screening can detect early polyps and guide removal to prevent development of colon cancer, results show only about 50% of members 50-80 years of age enrolled in commercial and Medicare plans are

receiving CRC screening. Furthermore, these results are consistent with published literature on colonoscopy as a favored screening method and persistent low rates of CRC screening.

Implications for Policy, Delivery, or Practice: These results highlight the need to continue to monitor and drive improved CRC screening in at risk populations. It illustrates claims/encounter data is reliable to capture screening.

Primary Funding Source: No Funding

●Development of an Outcomes-Based Pay-for-Performance Process for Outpatient Physical and Occupational Therapy

Dennis Hart, Ph.D., PT, Jerome B. Connolly, PT, CAE, Kathy Scott, PT, Mark W. Werneke, PT, MS, Dip. MDT

Presented By: Dennis Hart, Ph.D., PT, Director, Consulting and Research, Focus On Therapeutic Outcomes, Inc., 551 Yopps Cove Road, White Stone, VA 22578; Tel: (804) 436-9727; Fax: (804) 436-9328; Email: dsailhart@verizon.net

Research Objective: The pay-for-performance concept is rapidly influencing payers and legislators, but most pay-for-performance plans are based on process, not outcome. We describe development of an outcomes-based risk-adjusted pay-for-performance (value-based purchasing) model for patients receiving outpatient physical or occupational therapy. The model compares two variables: functional status (FS) change (effectiveness) and number of treatment visits (efficiency). We describe the predictive validity of the model and development of a value-based purchasing payment algorithm designed to encourage clinicians to be more effective and efficient. We compare differences in reimbursement for providing therapy using the algorithm compared to a fee-for-service plan.

Study Design: Secondary analysis of retrospective data.

Population Studied: Data from 189,088 patients (50 yr, SD 16, min 18, max 102, 39% male) in the Focus On Therapeutic Outcomes, Inc. data set from 552 outpatient clinics in 40 states (2000-2003) were analyzed. Most had orthopedic conditions. Patients entered FS data at rehabilitation intake and discharge using self-report surveys (change=discharge-intake). Therapists entered administrative data at discharge.

Principal Findings: Patients were randomly separated into two equal samples: developmental, testing. In each sample, discharge FS was analyzed using multivariate models controlling for 12 independent variables including: condition severity, age, symptom acuity, impairment category, payer source (including Medicare), and region of the country. The model controlled 35% of data variance. Beta coefficients of the independent variables were similar ($P>.05$) between samples, supporting cross-validation of the model. Beta coefficients of the developmental sample were used to predict discharge FS in the testing sample. A predictive ratio was estimated (1.09) using the testing sample by dividing predicted discharge FS by actual discharge FS, which supported predictive validity of the model. A value-purchasing payment algorithm was developed using the three most important independent variables from the regression analyses (condition severity, age, symptom acuity) plus impairment category (patient diagnosis). To facilitate implementation, patients were partitioned into 396 risk-adjusted cells (4 severity, 3 age, 3 acuity, 11 impairment categories). In each cell, patients were placed into 9 payment

groups using above expected, expected, and below expected FS change and visits. We developed a level of payment for each group that collectively were designed to encourage clinicians to produce good outcomes efficiently: bonuses, if more than expected effectiveness was reported using fewer than expected treatment visits; penalties, if less than expected effectiveness and more than expected visits. Using the payment algorithm in a real data simulation on the sample, a 12% reduction in payment was estimated had providers been reimbursed using the algorithm compared to fee-for-service.

Conclusions: Results support a risk-adjusted pay-for-performance model based on a measure of effectiveness could be developed for patients receiving outpatient therapy. The predictive validity of the model was supported. A simulation using the model reduced reimbursement.

Implications for Policy, Delivery, or Practice: A value-based purchasing process that is based on clinical outcomes and independent of treatment or provider may encourage providers to use evidence-based rehabilitation efficiently and may allow payers to redistribute resources to patients who are benefiting from treatment. The value-based purchasing process could be an alternative to the annual per beneficiary therapy caps for outpatient rehabilitation under Medicare Part B.

Primary Funding Source: No Funding

●Diagnosis of Acute Appendicitis by Using US and CT Imaging: Evaluation by Meta-Analysis

Chun-Bae Kim, Ph.D., Kun-Il Kim, Ph.D., In Soo Hong, Ph.D., Myoung Soo Kim, Ph.D., Hae Sook Sohn, Ph.D., Jean M. Abraham, Ph.D.

Presented By: Chun-Bae Kim, Ph.D., Associate Professor, Department of Preventive Medicine, Yonsei University Wonju College of Medicine, 162 Ilsan-dong, Wonju City Gangwon-do, 220-701; Tel: 82-33-741-0344; Fax: 82-33-747-0409; Email: kimcb@wonju.yonsei.ac.kr

Research Objective: To review the usefulness of ultrasonography (US) and computed tomography (CT) for the diagnosis of appendicitis and to evaluate the diagnostic accuracy of US and CT according to study characteristics.

Study Design: The Hasselblad method in meta-analysis was used to obtain the combined estimates of sensitivity and specificity for the outcomes of US and CT.

Population Studied: The relevant all-English languages articles published between 1981 and 2004 were included in this study if the patients had clinical symptoms of acute appendicitis. The histopathologic findings were the reference standard and the data were presented for ascertaining performance of US and CT in diagnosing appendicitis. Articles were excluded if patients had no sonographic or CT imaging signs of appendicitis. Two reviewer teams independently extracted the data on study characteristics.

Principal Findings: Nineteen articles (5,599 patients) fulfilled all inclusion criteria. The estimate of d in US and CT was 1.5782 [95% confidence interval (CI): 1.4583, 1.6982] and 1.9351 (95% CI: 1.7607, 2.1096) by a fixed effects model (p-value>0.05), respectively. Overall sensitivity and specificity of US were 72.3%, and 87.5%, respectively. Also, overall sensitivity and specificity of CT were 90.8%, and 88.8%, respectively. According to subgroup meta-analysis by study characteristics, the d estimate of dominantly younger age,

female, non-USA countries, and prospective study design group for US was 1.8368, 1.2660, 2.1111, and 1.7518, respectively. Also, the d value for CT done by same characteristics was 2.3703, 1.9252, 2.2609 and 2.0407, respectively.

Conclusions: This evidence suggests that CT may be accurate in diagnosis of acute appendicitis than US, especially when symptomatic patients are younger age and female.

Implications for Policy, Delivery, or Practice: This result can be used as evidence for comparing usefulness of new diagnostic technology on the management of appendicitis patient.

Primary Funding Source: Other Government

●An Evidence-Based Approach to Decreasing Patient Falls in a Major Cancer Center

Nancy Kline, Ph.D., RN, CPNP, FAAN, Patricia Brosnan, RN, Tara Cleary, RN, MA GNP-C, Maria Costino, RN, Wayne Quashi, MSN, RN, Bridgette Thom, MS

Presented By: Nancy Kline, Ph.D., RN, CPNP, FAAN, Director, Center for Evidence-Based Practice and Research, Nursing, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, M2102D, New York, NY 10021; Tel: (212) 639-6647; Fax: (212) 717-3178; Email: klinien@mskcc.org

Research Objective: This project takes an evidence-based, interdisciplinary approach to decreasing the number of patient falls and injuries attributable to those falls. The falls prevention team was comprised of clinicians and researchers from the nursing, rehabilitation, medicine, pharmacy, and quality assessment divisions. Nursing representation included nurse educators, geriatric and neurological clinical nurse specialists, nurse practitioners, and clinical nurses from both inpatient and ambulatory settings.

Study Design: After thorough review of current literature and existing falls-risk assessment instruments, we conducted a case-control study to identify variables that distinguished patients who fell from those who did not. This initial list of variables was developed from factors commonly named in the literature as well as those identified in clinical practice.

Population Studied: Each patient at our institution, a comprehensive cancer center, who fell in the first quarter of 2005 was matched with two patients of a similar age from his/her floor, indicating a similar diagnosis. Patients had a variety of cancer diagnoses, but all were on inpatient units.

Principal Findings: The charts of 219 patients, 73 cases and 146 controls, were reviewed, and statistical significance, through chi-square testing, was found in five variables: at .10, cognitive impairment and motor deficits; at .05, sensory deficits and medication count; and, at .000, history of falls. Odd-ratios calculations produced similar findings.

Conclusions: We developed a falls-risk assessment instrument based on the case control findings and previously developed instruments in the literature. Included in our instrument was an assessment of the patient's ability to complete basic activities of daily living, which has been shown to be of use in predicting patients at risk for falls in a previous JCAHO case-control study. Interrater reliability of the instrument was 84 percent.

Implications for Policy, Delivery, or Practice: A corresponding falls-risk stratification system, which assigns appropriate interventions based on a patient's score, was

developed and includes two risk categories, regular and high, and specific interventions for each risk variable measured. Among others, interventions at the regular-risk level include ensuring basic patient and environmental safety precautions are met and educating the patient and family about prevention of falls. In addition to the regular-risk interventions, high-risk patients are observed more closely and supervised and assisted while toileting and transferring. High-risk status is flagged on the medical record with red stars and on the patient with red wrist bands. High-risk patients are assessed for the need for physical or occupational therapy consultation, a 24-hour sitter, and relocation to a room with the best visual access to the nursing station. Interventions to address patient-specific risk variables--including mental status, elimination, sensory and motor deficits, age, fall history, mobility, and medications--are outlined. The instrument is currently being pilot-tested on four inpatient units: neurology, leukemia/lymphoma, head and neck, and hepatobiliary and gastric mixed tumor. Patients are assessed for falls risk at admission and twice daily during hospitalization. If a patient falls, a team responds to review the risk score and additional environmental and patient variables present at the time of the fall. Data are being collected prospectively, and the instrument will be modified as needed.

Primary Funding Source: No Funding

•Survey of Physician Attitudes about and Response to Performance Incentives Being Implemented by Minnesota Purchasers and Payers

David Knutson, MS, Jinnet Fowles, Ph.D., Michael Callahan, MS, Dan McGaughlin, M.B.A.

Presented By: David Knutson, MS, Director, Park Nicollet Institute, Health System Studies, 3800 Park Nicollet Blvd., Minneapolis, MN 55416; Tel: (952) 993-3287; Fax: (953) 993-3741; Email: knutsd@parknicollet.com

Research Objective: To determine physician practice attitudes about performance incentives being implemented nationally and locally by purchasers and payers. These include pay-for-performance, public reporting of physician practice performance, and consumer price tiering of physician practices.

Study Design: A web-based survey, followed by a mailed survey of non-respondents, was conducted of all individual members of the MMGMA who concurrently held management responsibilities for a physician practice. The sample was subdivided into practices that provided primary care. Where there were multiple respondents from a practice, the highest-level administrator was selected. By this method, we retained respondents from 78 physician practices providing primary care. The practice sizes ranged from 3 to 1,000 physicians with a median of 21.

Population Studied: Individual members (563) of the Minnesota Medical Group Management Association (MMGMA) were surveyed.

Principal Findings: Data collection was completed in November, 2005. Data cleaning and descriptive statistical analysis have been completed. While we can report selected descriptive statistics now, the paper will include results from bi-variate and multivariate analyses being conducted. The survey explored a wide range of topics related to provider incentive approaches, methods, implementation, impact on

patients, and internal response of practices. Due to size limitations, only a few selected findings can be reported here. About 63% of practices reported having pay-for-performance contracts. Of those, 65% reported that the performance revenue represented less than 2% of total revenue, while only 4% reported greater than 10%. About two-thirds of practices agreed or strongly agreed that pay-for-performance would increase use of evidence-based medicine. Several methodological concerns and cross-payer inconsistencies were also reported. For example, 61% reported that current performance contracts do not adequately account for patient case-mix. Regarding public reporting that compares practice performance, questions were asked about the policy in general and about two prominent examples in Minnesota. Responding to a statement that the MN Community Measurement Initiative quality measures are easily understood by patients, 47% disagreed or strongly disagreed. The rest reported that they did not know. Also, responding to a statement that public reporting would produce changes in market share, 56% reported they were uncertain, while 27% disagreed. We also asked about network tiering in general, and specifically of two examples in Minnesota.

Conclusions: Minnesota physician practices seem to support the potential of pay-for-performance, but have some concern about implementation issues. They are skeptical about the methods of public reporting and are unsure of its impact on patient demand. They are more aware and engaged about network tiering, and are concerned about methods and how much control they have over large cost drivers such as specialists and hospitals. There is in general a high degree of uncertainty as yet about the methods and impact of recently emerging physician practice performance incentives in Minnesota, but the majority believes the trend is moving toward producing the intended outcomes.

Implications for Policy, Delivery, or Practice: As multiple public and private purchasers and health plans increasingly introduce a wide array of physician incentive programs to reward quality, we need to assess how these multiple and simultaneous approaches are perceived and responded to by physician practices. Such knowledge can further the alignment of purchasing strategies to accelerate the translation of quality goals into practice at the community level.

Primary Funding Source: AHRQ

•The Influence of Agency Characteristics on Home Health Care Outcomes

Elizabeth Madigan, Ph.D. RN, Catherine Vanderboom, Ph.D., RN

Presented By: Elizabeth Madigan, Ph.D. RN, Associate Professor, Frances Payne Bolton School of Nursing, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106-4904; Tel: 216-368-8532; Fax: 216-368-3542; Email: elizabeth.madigan@case.edu

Research Objective: Determine whether agency characteristics (type, location, and years certified) influence risk adjusted outcomes in a national population of home health care agencies

Study Design: The Center for Medicare and Medicaid Services reports risk adjusted data on all Medicare certified home health care agencies for 10 outcomes: 4 for improvement in functional status items (bathing, transferring,

walking, management of oral medications); 3 for improvement in clinical outcomes (urinary incontinence, pain interfering with activity and dyspnea); and 3 health service utilization outcomes (unplanned hospitalization, requiring urgent care, and remaining at home following home health care discharge). CMS data include agency provider number, years in operation, and type of agency (governmental, for-profit, and non-profit). These agency-level outcomes were linked via provider number zip code to the Rural-Urban Commuting Area codes to determine geographic location (urban, town or rural).

Population Studied: Data were obtained in January 2006 from the CMS web-based public reporting system for home health care, Home Health Compare, and linked with RUCA files from the University of Washington Rural Health Research Center. The population consists of 7735 agencies.

Principal Findings: The risk adjusted outcomes are reported such that higher numbers for functional and clinical items indicate better agency quality performance while higher numbers for utilization of services indicates worse agency quality performance. Mean improvement in functional status ranged from 39% for walking to 60% for bathing and for the clinical items ranged from 55% (urinary incontinence) to 61% (improved pain during activity). Risk adjusted hospitalization was 30.8%, urgent care use was 28.5% and 62.6% of patients remained at home following home health care. Longer years in service were associated with worse performance in functional and clinical risk-adjusted outcomes ($r = -.16$ to $-.25$) but better performance for hospitalization ($r = -.026$) and requiring urgent care ($-.28$). Urban agencies had better outcomes than town and rural agencies for all functional and clinical outcomes but also higher percentages of hospitalization (32% vs. 27 and 28%, respectively) and urgent care use (30% vs. 23% and 25% respectively). Town and rural agencies were similar for most measures. Similarly, for-profit agencies (compared to governmental and non-profits) had better functional status and clinical outcomes but worse performance for hospitalization (35% versus 27 and 23%, respectively) and urgent care use (33% versus 23 and 20%, respectively). In multivariate analysis, the variance explained for all regressions were small, $< 10\%$, and years in service had the highest standardized regression coefficient for each outcome (betas $\sim .25$).

Conclusions: There appears to be a trade-off where agencies whose risk adjusted outcomes are better for functional and clinical performance have worse outcomes for hospitalization and urgent care use. Years in service is the strongest indicator of risk adjusted outcome with type and location having relatively little influence, although the overall explained variance is small.

Implications for Policy, Delivery, or Practice: Policy changes to encourage growth in specific agency types or locations may not improve overall outcomes in Medicare home health care beneficiaries.

Primary Funding Source: Ohio Board of Regents Research Infrastructure Grants to Case Western Reserve University

●Consumer and Physician Interpretations of Health Care Quality: Integrating Technical and Experience Quality Indicators

Ann Mirabito, Ph.D. Candidate, Jelena Spanjol, Ph.D., Leonard Berry, Ph.D.

Presented By: Ann Mirabito, Ph.D. Candidate, Doctoral candidate, Dept of Marketing, Mays Business School, Texas A&M University, 4112 TAMU, College Station, TX 77843-4112; Tel: 979-845-4525; Fax: 979-862-2811; Email: amirabito@tamu.edu

Research Objective: Health care quality encompasses both technical and experience components. This research explores the similarities and differences in the way consumers and physicians integrate technical and experience quality cues to form overall quality evaluations. The literature suggests that consumers and physicians evaluate the two components differently and that consumer preferences for quality are unstable, i.e. the expressed preference is shaped by time, context and elicitation method. In the first phase, we develop a framework for understanding and measuring health care experience quality. In the second phase, we investigate the similarities and differences in the ways consumers and physicians integrate their interpretation of technical and experience quality cues in forming overall health care quality judgments and preferences.

Study Design: To develop the framework for understanding experience quality, we conducted an extensive review of the health care experience quality literature, examining medical, business, and psychology databases. Both conceptual and empirical studies were reviewed. A similar review was conducted of experience quality attitudes for other credence services, that is, services for which quality is difficult to judge even after the service has been provided. Then, findings were synthesized and a framework for evaluating health care experience quality was developed.

Population Studied: Conceptual and empirical studies of experience quality in health care and in other credence services including law and financial services, as reported in medical, business and psychology databases.

Principal Findings: We found evidence of preference reversal when consumers were asked to choose between technical and experience quality. However, preferences for the experience quality dimension appear to be stable. We also found that the structure-process-outcome framework used to evaluate technical quality does not fit experience quality.

Conclusions: Drawing on the literature in services quality and empirical studies of consumer and physician attitudes toward experience quality, we establish a framework for evaluating experience quality that parallels the structure-process-outcome framework for evaluating technical quality. We also conclude that the problem-solving literature, particularly the literature dealing with expert-novice categorization processes, sheds light on how people integrate technical and experience quality. Preliminary findings from the second phase of the project, involving the similarities and differences in the ways consumers and physicians integrate technical and experience quality, will be presented.

Implications for Policy, Delivery, or Practice: Consumers are expected to become more proactive in seeking out health care quality. But there is concern that consumers appear to underuse the existing, albeit limited, quality assessment tools

and that, from a normative perspective, consumers may be relying on suboptimal quality indicators. This research provides a framework for understanding how consumers use experience quality cues to infer technical quality and how they make technical and experience quality tradeoffs as needed. This research also bridges the gap between physician and consumer attitudes toward quality by showing how the two stakeholders interpret quality.

Primary Funding Source: No Funding

●**Examining the Impact of Nursing Home Resident and Facility Characteristics on the Reliability of Multiple MDS Measures**

Vincent Mor, Ph.D, Ning Wu, Ph.D, Jason Roy, Ph.D.

Presented By: Vincent Mor, Ph.D, Professor, Community Health, Brown University, 167 Angell St, Providence, RI 02912; Tel: (401) 863-2959; Fax: (401) 863-3492; Email: Vincent_Mor@brown.edu

Research Objective: In the nursing home (NH) setting, the Minimum Data Set (MDS) is used as the basis for prospective payment and quality of care monitoring. With the promotion of Pay for Performance, the Center for Medicare and Medicaid Services (CMS) is testing the linking between payment and quality of care, which can be partially measured by the quality indicators derived from the MDS. Measurement bias in the data files will have substantial impact on all the applications. We examined variation in the coding of multiple MDS items among NHs and the impact of resident and facility characteristics on coding accuracy.

Study Design: We used the data from a reliability study. Research nurses and facility staff independently conducted MDS assessments on 5758 residents from 209 facilities in 6 states. Both research nurses and NH staff were instructed to follow the assessment protocol in Resident Assessment Instrument User's manual. Because research nurses were uniformly trained and experienced, their assessments were considered the "gold standard," whereas assessments completed by facility staff were more likely to vary due to variation in training, experience or facility specialty, etc. We examined the coding differences between facility and research nurses of eight MDS-derived measures in various health domains. These measures are included in the calculation of quality indicators that CMS releases to public. Multilevel models were built to examine the impact of resident and facility characteristics as well as the location of the facility (state) on coding differences.

Population Studied: Nursing home residents

Principal Findings: Across all the measures there was substantial within- and between-facility variation in the coding differences between facility and research nurses. Resident characteristics were unable to explain the variations, whereas facilities characteristics explained 5 to 20%, and indicators of NH location (state) further explained 15 to 35% of the variation. We also observed a generalized state effect that tended (although not uniformly) to be consistent across measures. Compared to facilities in Tennessee, facilities in California overrated limitations in Activities of Daily Living, cognitive impairment, stage of pressure ulcer and pain to a larger extent; facilities in Missouri underrated delirium, pain, and bowel and bladder incontinence. The results suggest that there are some state-level factors, e.g., state policies or

training programs, that have an impact on measurement error. States that over-identify problems in multiple measures also tend to have high quality indicators, and vice versa.

Conclusions: Our results suggest that state associations and training organizations may impose a systematic shift in the interpretation of how to apply the assessment criteria. Unless we adjust for or eliminate the substantial variation in measurement across states, comparison of quality of care and payment based on performance will be adversely impacted.

Implications for Policy, Delivery, or Practice: We call for an experiment to test whether state differences are attributable to training or to inconsistency in how state "experts" charged with answering providers' queries respond to specific questions.

Primary Funding Source: No Funding

●**Do Different Approaches to Risk-Adjustment Affect Home Health Agency Quality Rankings?**

Timothy Peng, Ph.D., Christopher Murtaugh, Ph.D., Hakan Aykan, Ph.D., Gil Maduro, Ph.D.

Presented By: Timothy Peng, Ph.D., Senior Research Associate, Center for Home Care Policy and Research, Visiting Nurse Service of New York, 5 Penn Plaza, New York, NY 10001; Tel: 212-609-5765; Fax: 212-290-3756; Email: tpeng@vnsny.org

Research Objective: There are currently 31 home health quality measures risk adjusted by the Centers for Medicare and Medicaid Services (CMS) and reported to agencies in the Outcome-Based Quality Improvement program (OBQI). A subset of these measures is publicly reported through Home Health Compare (HHC). A data-driven "stepwise" approach is used to risk adjust the quality indicators, with a separate set of risk factors included in the logistic regression models for each measure. The stepwise approach relies on a large number of risk adjustors and its complexity may present barriers to providers who may wish to engage in local quality improvement efforts. This project tests the use of a theory and evidence-based alternative approach to risk adjustment, examining effects on predicted outcomes and relative performance rankings of home health agencies on each of the 31 OBQI indicators, using the same source data as CMS to construct the models. The theory-driven approach follows lezzoni's recommendations for developing risk adjustment models for health care outcomes, taking into account current research in the home health setting.

Study Design: The data analyzed in this project were obtained from the CMS contractor at the University of Colorado. They drew the data from the OASIS National Repository at CMS to create discrete episodes of home health care during calendar year 2001. Alternative models were estimated sequentially after replicating the current risk adjustment models. The first model was limited to the admission (or baseline) value of the outcome indicator and a core set of risk adjusters. Subsequent models included a small number of outcome-specific risk adjusters. Following development of a final set of alternative risk adjustment models, an agency-level analysis was conducted to determine the impact on agencies' quality ratings.

Population Studied: All episodes of Medicare home health beginning and ending within calendar year 2001.

Principal Findings: Across the 31 OBQI outcomes, the alternative models produced similar results. Alternative models generally were slightly lower in explanatory power, as expected, when compared to models estimated using the stepwise approach. Overall agency-level quality ratings were very similar for the alternative and the current “stepwise” approach (absolute mean differences ranging from -0.22 to +0.33 percentage points). Spearman’s rank correlations between the alternative and the current models also demonstrated very close correspondence in relative Agency rankings, ranging from .93 to .98, with the exception of Improvement in Ambulation and Improvement in UTI models (.89 and .90, respectively).

Conclusions: The results suggest that using a theory and evidence-based approach to risk-adjustment models for the OBQI indicators is unlikely to have a substantial effect on the quality ratings of the majority of home health agencies. Selection of either approach yields similar findings with respect to agency-level performance rankings on home health quality measures, with respect to individual outcomes within a given agency (mean differences) or relative to the performance of other agencies (ranking).

Implications for Policy, Delivery, or Practice: Given that both the current “stepwise” and the theory- and evidence-based alternative presented here produce very similar results, payers, providers, and policymakers need to consider the tradeoffs of using either approach to risk adjust measures of home health quality and performance. The current “stepwise” framework provides the greatest ability to maximize the explanatory power of the risk adjustment models. However, it also runs the risk of changing over time, requires collection of a large pool of risk-adjusters, and lacks transparency since the actual risk-adjusters remain unknown until after complete national data have been collected and fully analyzed. The alternative theory-driven approach has the potential to provide consistency and stability over time. Its relative simplicity affords agencies greater ability to engage in risk-adjustment of home health outcomes in their own quality improvement efforts.

Primary Funding Source: DHHS, ASPE

●Use of the No-Blame Box in Reaching Consensus on the Cause of Error in Cancer Diagnosis

Stephen Raab, M.D., Dana M. Grzybicki, M.D., Ph.D., Chad H. Stone, M.D., Eva M. Wojcik, M.D., Kim R. Geisinger, M.D., Laila Dahmouh, M.D.

Presented By: Stephen Raab, M.D., Director of Pathology Quality and Healthcare Research, Pathology, University of Pittsburgh, 5230 Centre Avenue, Pittsburgh, PA 15232; Tel: (412) 623-5950; Fax: (412) 623-4014; Email: raabss@upmc.edu

Research Objective: Physicians do not agree on the cause of cancer diagnostic errors, which contributes to the failure to reduce these errors. We wanted to determine if the No-Blame Box method of root cause analysis results in better interobserver assessment of cancer diagnostic error cause.

Study Design: We designed a visual tool, known as the No-Blame Box, that allows for individuals to perform error root cause analysis by evaluating several error metrics simultaneously. The No-Blame Box allows for individuals to focus on multiple inter-related processes, rather than healthcare professionals, when adjudicating the cause of error.

Using the No-Blame Box method of root cause analysis, 5 pathologists blindly assessed the cause of error in 40 patient specimens procured from 6 hospitals. The method consisted of separately assessing the specimen quality and the presence of tumor, prior to assigning the error cause as a failure in tissue procurement or a mistake in diagnosis or both. Interobserver variability of assessment of error cause was measured using the kappa statistic.

Population Studied: Forty patients who had errors in anatomic pathology tests obtained to determine the presence or absence of lung cancer.

Principal Findings: The kappa statistic of all pathologist pairs in assessing error cause was greater than 0.400, range: 0.400 to 0.796, indicating acceptable to excellent agreement. A statistically significant association existed between pathologist and assignment of specimen interpretability, $P = 0.044$. In no case was there unanimous agreement that the specimen sample was excellent and tumor was present. All pathologists agreed that tumor was present in 43% of the cases, but at least one pathologist thought that the specimen quality markedly limited interpretability. Errors were a result of process breakdown and failure to procure a uniformly interpretable sample.

Conclusions: The No-Blame Box method of root cause analysis results in significantly better agreement in determining the cause of errors in cancer diagnosis. Standardizing this process on a national level would result in more effective error reduction in diagnostic testing for cancer. The No-Blame Box may be used as a visual tool to adjudicate causes of error and as a means to show the interplay between different causes.

Implications for Policy, Delivery, or Practice: Agreement on cause of cancer diagnostic errors allows for the design of error reduction initiatives. Currently, Congress mandates that labs perform error cause assessment, and the No-Blame Box may be used to move beyond error reporting to error reduction on a national scale.

Primary Funding Source: AHRQ

●Increasing Generic Prescribing Rates: The Impact of a Generic Rx Incentive Program For Physicians

Tanja Rapus Benton, Ph.D., Soyol Momin, MS, M.B.A., Kent Barnes, Beverly Franklin-Thompson, PharmD, Pete Van Valkinburgh, M.B.A., Steve Coulter, M.D.

Presented By: Tanja Rapus Benton, Ph.D., Sr. Bio-Statistical Research Analyst, Health Services Research, BlueCross BlueShield of Tennessee, 801 Pine St., Chattanooga, TN 37402; Tel: (423) 535-6924; Fax: (423) 535-5100; Email: Tanja_Benton@BCBST.com

Research Objective: The objective of the program was to increase the prescription of less costly generically available drugs by providing incentive to physicians. By setting annual generic prescribing goals, and sharing the resultant savings with the provider group the trend of rising prescription costs may be slowed.

Study Design: Generic prescribing goals were set by comparing current generic prescription rates for several provider groups with the current trended health plan average. The incentive amount was based on estimated savings resulting from an increase in generic prescribing. Actuarial estimates indicate that prescription drug costs are reduced 1%

for every 1% increase in generic utilization. Measurement of the generic prescribing rate was based on a rolling 12 months of retail pharmacy claims data beginning January 2004. Feedback was provided quarterly for each provider group as well as each individual provider. The resulting shared savings amount was then applied to the groups' fee schedules after the final annual performance evaluation.

Population Studied: Three distinct provider groups participating in the commercially insured population administered by a large single-state health insurance carrier in the southeastern U.S. were selected for the program: a primary care IPA (n = 49); a university-based multi-specialty group (n = 100); and a university family practice group (n = 27). The average generic prescribing rate for these groups was 39%, which was lower than the health plan average of 42%.

Principal Findings: The findings show that the generic prescribing rate increased for all three provider groups after the first year of the program. The average generic prescribing rate reached 47%, compared to a pre-program rate of only 39%, and matched the health plan average.

Conclusions: An incentive-based approach to generating increases in generic prescription rates appears to be an effective strategy to mitigate rising prescription costs.

Implications for Policy, Delivery, or Practice: This study has several important implications for how incentive-based plans should be structured and evaluated. First, how the incentive is structured and delivered may be an important factor in producing changes in performance. A more immediate and direct incentive, such as a bonus payment, may have greater impact than a percentage increase in fee schedule delivered on a yearly basis. Second, the data used to measure the generic prescribing rate were based on retail pharmacy claims, thus it is difficult to determine whether increases in the generic prescribing rate can be attributed to an increase in scripts written for generic drugs at the physician level or how prescriptions are filled at the pharmacy level. Rate increases may reflect an increase in generic substitutions when prescriptions are filled. Third, in some cases there may not be generic substitutions available. A better measure of the generic prescribing rate would be based only on drugs for which generic substitutions are available. Consequently, generic incentive plans may be more viable for some specialties. By improving the design of incentive plans as well as the methodology used for assessing success, more effective incentive-based programs can be implemented.

Primary Funding Source: BlueCross BlueShield of Tennessee

● Implementation of Patient Safety Measures: Trends over Time in the Veterans Health Administration (VA)

Amy Rosen, Ph.D., Shibe Zhao, M.P.H., Peter Rivard, M.H.S.A., Susan Loveland, M.A.T., Anne Elixhauser, Ph.D., Patrick Romano, M.D., M.P.H.

Presented By: Amy Rosen, Ph.D., Senior Research Scientist, Center for Health Quality, Outcomes and Economic Research, 200 Springs Road, Bedford, MA 01730; Tel: (781)-687-2960; Fax: (781)-687-3106; Email: akrosen@bu.edu

Research Objective: The Patient Safety Indicators (PSIs), developed by the Agency for Healthcare Research and Quality (AHRQ), were designed to capture those instances representing potentially preventable adverse events that compromise patient safety in the inpatient setting, such as

complications following surgeries, other procedures, and medical care. Although the PSIs are considered indicators, not definitive measures of patient safety, they are useful screening tools for highlighting areas in which quality should be further investigated and providing useful benchmarks for tracking progress. The purpose of this study is to examine trends in PSIs over time in the VA, a national leader in patient safety. Our specific objectives are to: 1) provide a descriptive analysis of the incidence of PSI events over time (2001-2004); 2) examine whether trends in national PSI rates are significant; and 3) assess whether hospital characteristics and hospitals' risk-adjusted rates in 2001 predict future hospital performance (2004 risk-adjusted rates).

Study Design: We determined unadjusted and risk-adjusted rates of potential PSIs by applying the PSI software to VA inpatient administrative files. Generalized linear models (GLM) were used to assess trends in risk-adjusted rates over time, and to examine whether hospital characteristics in 2001 (teaching status, number of beds, location, volume, and mean quality improvement (QI) implementation score) and hospitals' 2001 risk-adjusted performance were predictive of hospitals' risk-adjusted future performance.

Population Studied: All veterans discharged from acute care hospitals from 2001-2004. We merged all hospital discharge summary files into one aggregated file containing hospitalizations with and without PSIs during this period. A total of 108 hospitals were present in all years.

Principal Findings: The most frequent PSI events across all years were failure to rescue, decubitus ulcer, accidental puncture/laceration, and pulmonary embolism or deep vein thrombosis (PE/DVT). Of the 14 PSIs examined, we found a significant increasing trend over time ($p < 0.05$) for iatrogenic pneumothorax, and a significant decreasing trend ($p < 0.01$) for failure to rescue (deaths following complications).

Postoperative PE/DVT and infection due to medical care had borderline significant increasing trends. Rates of high-frequency medical/surgical complications and surgical/technical complications were stable across years. After accounting for patient and hospital characteristics, hospitals' 2001 risk-adjusted PSI rates were the most important predictors of future performance for 8 PSIs, suggesting the consistency of patient care practices and hospital coding/documentation practices over time.

Conclusions: Similar patterns of PSI event rates across years provides further evidence of the reliability of the PSIs as indicators of potential patient safety events. We observed some increase in selected PSI events over time, suggesting that the VA focus attention on these PSIs, through QI initiatives. Specific consideration should be given to those PSIs that are potentially preventable, relatively frequent, and have established processes of care, such as postoperative PE/DVT.

Implications for Policy, Delivery, or Practice: Although the PSIs are useful tools for tracking and monitoring patient safety events, future research should investigate whether trends in PSIs reflect changes in quality of care or increased attention to documenting patient safety events.

Primary Funding Source: VA

●Relationship of Individual Personnel Characteristics to Safety Culture in U.S. Hospitals

Sara Singer, M.B.A., Laurence Baker, Ph.D., Sara Singer, M.B.A., Alyson Falwell, M.P.H.

Presented By: Sara Singer, M.B.A., Doctoral Candidate, Ph.D. Program in Health Policy, Harvard University, Sherman Hall 303-1, Boston, MA 02163; Tel: (617) 495-5047; Fax: (617) 496-4397; Email: ssinger@hbs.edu

Research Objective: The presence of a hospital culture that emphasizes patient safety is increasingly recognized as a key to improving patient safety. Perceived and actual safety culture may vary from one part of an institution to another, with potential implications for efforts to improve safety cultures. Though some research has examined perceptions of safety culture among personnel with varying job characteristics, these relationships are not well understood. This AHRQ-supported research sought to understand relationships between perceptions that individuals working in hospitals have about safety culture and their management status, clinical status, and work area.

Study Design: We performed a cross sectional comparison of the relationship between individual characteristics and their perceptions about safety culture in their hospital. Safety culture measures are derived from the Patient Safety Climate in Healthcare Organizations survey, administered to personnel at nationwide sample of hospitals in 2004. Results from the survey provide an overall measure of each respondent's perception of the strength of safety culture at their institution, and sub-scale measures specific to several individual aspects of safety culture. A response rate of 52.7% and psychometric analysis of response patterns support the validity of the data. The survey also gathered data on respondents' management status, clinical status, work area, and other demographic characteristics. Regression analysis related culture measures to individual characteristics of interest, controlling for the respondent's age, gender, job tenure, as well as the size, tax, teaching and urban status of their hospital.

Population Studied: 16,842 individuals from a stratified random sample of 105 hospitals, representing all 4 regions of the U.S. and 3 size categories. In each hospital, 100% of senior managers, defined as department head or above, 100% of active physicians, and 10% of other staff were surveyed.

Principal Findings: A. Senior managers have much more positive perceptions of safety culture than supervisors and front-line workers, overall and in each of the individual sub-scales examined. B. Individuals who work in high-hazard units have more negative perceptions of safety culture than other individuals. C. Nurses have more negative perceptions of safety culture than other clinicians and non-clinicians, particularly in some areas. For example, they are more likely to perceive gaps in knowledge about safety issues between front line workers and senior managers, and problems with support and rewards for safe care.

Conclusions: Consistent with prior research, we found that senior managers viewed hospital safety culture more favorably than other personnel. Personnel working in high-hazard areas viewed safety culture as more problematic, which is concerning given the high potential payoff to very strong safety cultures in these units. Variations in perception of safety culture between nurses, physicians, and others may be

related to particular demands of different roles, but could also signal challenges in achieving a strong uniform safety culture in complex institutions.

Implications for Policy, Delivery, or Practice: Efforts to achieve strong uniform safety cultures will face significant challenges, and may require activities tailored to the different challenges facing different types of personnel. Activities that can reduce differences in perceptions across levels within organizations may be particularly valuable, as may activities that focus on culture in high-hazard units.

Primary Funding Source: AHRQ

●Identifying Close Calls: Interpretations of a Near Miss

Michal Tamuz, Ph.D., Eric J. Thomas, M.D., M.P.H.

Presented By: Michal Tamuz, Ph.D., Associate Professor, The Center for Health Services Research, The University of Tennessee Health Science Center, 66 North Pauline, Suite 463, Memphis, TN 38163; Tel: 901-448-3716; Fax: 901-448-8009; Email: mtamuz@utm.edu

Research Objective: To understand how healthcare professionals perceive and interpret a near miss

Study Design: Healthcare quality experts believe that near miss reporting in hospitals will help identify opportunities to improve systems of care. However, near misses and close calls are ambiguous events. A near miss can be an opportunity to identify flaws in systems prior to an adverse event, or a near miss may be interpreted as a signal that systems are safe because no adverse event occurred. Using qualitative research methods, we analyzed interviews conducted as part of a larger research project on organizational learning. A semi-structured interview protocol was used in which randomly-selected healthcare professionals were asked to describe reporting systems and other programs designed to monitor and promote medication safety. Medication errors occur frequently in hospitals and intensive care units (ICUs), so we focused on the process of medication delivery in one ICU. Interviews with participants were audio-recorded and transcribed; the interview data were supplemented by field notes, document review, and observations of routine activities. Here we report our analysis of the participants' responses to a near miss vignette.

Population Studied: Hospital administrators and ICU healthcare providers in a tertiary care, teaching hospital (n=40)

Principal Findings: 1) Study participants varied in their perceptions of a near miss. They described the same event alternatively as evidence of a system breakdown (30%), a robust system (25%), and individual accountability 25% (other=10%, n/a=10%). 2) Perceptions differed by position. For example, 82% (n=11) administrators perceived the near miss as a system breakdown while only 11% (n=9) ICU staff nurses did. 3) We observed a pattern in which the lower the position in the hospital hierarchy, the more likely the individual would attribute the near miss to individual accountability rather than to system issues, ranging from 80% of the support staff (n=5) who attributed the near miss to individual accountability to 18% (n=11) of the hospital administrators. 4) Some practitioners (58%, n=12) who perceived the near miss as an indicator of a system breakdown spontaneously raised questions about medication safety practices, while no such questions were asked by those

who saw the near miss as reflecting individual accountability (n=10) or a robust system (n=10).

Conclusions: Providers and administrators differed in how they interpreted the meaning of a near miss. If near misses are believed to be signs of safe and effective systems, then providers may have little reason to report them. Alternatively, if near misses are seen as early warning signs of system flaws, the providers may be more likely to report them. Moreover, the potential of learning from a single near miss is reduced if it is interpreted in terms of individual accountability.

Implications for Policy, Delivery, or Practice: If confirmed by future studies, these findings have significance for the widespread adoption of near miss reporting systems in healthcare. Tacit near miss interpretations, if left unaddressed, may obstruct the effective implementation of patient safety reporting systems and organizational learning from near misses.

Primary Funding Source: Aetna Foundation's Quality Care Research Fund

●Hierarchical Bayesian Cost-Effectiveness Models for Cross-Sectional Provider Profiles

Justin Timbie, BA, Sharon-Lise Normand, Ph.D.

Presented By: Justin Timbie, BA, Doctoral Student, Ph.D. Program in Health Policy, Harvard University, 180 Longwood Ave, Suite 301, Boston, MA 02115; Tel: (617) 869-1371; Email: timbie@fas.harvard.edu

Research Objective: To combine measures of in-hospital survival and spending using a cost-effectiveness framework to obtain a summary measure of the value of hospital care for Acute Myocardial Infarction (AMI) for the purpose of profiling hospitals.

Study Design: Observational study using an all payer hospital discharge database. We estimate incremental costs and effectiveness for each hospital by taking differences between predicted and expected outcomes. Predicted outcomes are estimated using hierarchical models that allow hospitals to vary in their baseline quality and efficiency, while expected outcomes are estimated by fixing baseline quality and efficiency to be the same across hospitals, and allowing each estimate to be a function of case mix alone. Incremental outcomes reflect a hospital's actual performance relative to its expected performance, given its unique case mix. The pair of incremental outcomes are then combined to obtain estimates of incremental net benefit, and the performance of each hospital is compared using cost-effectiveness acceptability curves. The analysis assumes the perspective of a private payer.

Population Studied: The sample was comprised of all non-federal, acute care hospitals in the state of Massachusetts that treated at least one AMI patient during fiscal year 2003. Patient-level inclusion criteria included: age 18 years or older, primary discharge diagnosis of AMI (ICD-9 code 410.xx, excluding 410.x2), and a minimum length of stay of 2 days (if discharged alive). Patients with "Do Not Resuscitate" orders and patients who were discharged to hospice were excluded. Patients transferred between hospitals were included and their admissions were linked to create episodes of care provided that both admissions were contained in the database. A total of 69 hospitals and 11,128 patients were used in the analysis.

Principal Findings: After determining two threshold criteria—a willingness to pay threshold of \$5 million per life saved and a posterior probability of positive incremental net benefit exceeding 0.5—a total of 15 hospitals were designated as high value. Using the traditional approach of evaluating multiple outcomes—comparing the performance on both outcomes jointly—would lead to the selection of a different set of hospitals. The limitations of the method are discussed, including the choice of AMI in particular, use of charge data, and missing data on care provided in non-inpatient settings.

Conclusions: Cost-effectiveness methods can be applied to provider profiling to compare variation between hospitals in the tradeoff between survival and spending. The approach relies on a strong theoretical foundation and includes the most current methods from the profiling literature, including the use of Bayesian hierarchical models. The method is general enough to be applied to a wide range of conditions and their associated outcome measures.

Implications for Policy, Delivery, or Practice: Such a method can complement ongoing pay-for-performance or hospital tiering efforts, that rely predominantly on process measures and avoid measures of spending.

Primary Funding Source: Sloan Foundation for the Study of the Managed Care Industry

●Implementation Standards for Benchmarking Physicians on Resource Use (Efficiency)

Sally Turbyville, MA Economics, Dave Knutson, MS

Presented By: Sally Turbyville, MA Economics, Senior Health Care Analyst, Quality Measurement, NCQA, 2000 L Street, NW ste 500, Washington, DC 20036; Tel: 202-955-1756; Fax: 202-955-3599; Email: turbyville@ncqa.org

Research Objective: To evaluate the feasibility of the development of consensus-based implementation standards or parameters across and within proprietary software tools currently in the market for the purposes of regional and national profiling and benchmarking of physicians on cost efficiency or relative resource use. Common implementation standards would increase transparency and improve the usefulness of results for accountability.

Study Design: This work was based on peer-reviewed publications and expert panel advice. NCQA identified vendors that offer methods to calculate physician cost efficiency, have a stable market presence, and are considered reliable and sound approaches. NCQA contacted each vendor individually to describe the goal of this project and request participation. Each vendor responded to a request for information about their product and participated in at least one telephone follow-up call to discuss their responses and initial standards for implementation.

Population Studied: Five vendors representing six products participated. Three products characterize utilization at the patient level (person-level) and three grouped services into episodes and characterized utilization at the episode level.

Principal Findings: NCQA was able to identify implementation standards within each software tool and areas of common ground across tools. The standards addressed 3 critical steps in implementation: a) input data, b) methods used to estimate a patient's risk score and expected cost; and c) physician attribution. Input data refers to the types of data (e.g., pharmacy data), or the amount of data necessary to

reliably and accurately evaluate a physician's relative cost of care; input data can vary in ways that can produce substantial error, bias, or inconsistencies. The methods used to estimate a patient's risk score and expected cost refer to decision steps or procedures that affect the statistical reliability and fairness of the outcomes, such as outlier decision rules and sample size requirements. For example, a person-level tool may recommend truncating the top and bottom 5 percent most expensive patients. Lastly, attribution rules refer to which episodes or members and under what conditions they would be attributed to a physician. Physician attribution rules should attribute responsibility for costs to physicians that is commensurate with the degree of actual or desired influence of the physician on cost of care. We found that implementation standards for the input data were relatively simple for vendors to implement or change to follow a recommended standard. Methods used to estimate a patient risk or expected cost were less straight forward in finding standards among the tools. However, it was more common that we identified common standards among tools that use the same approach (person vs. episode). Vendors were able to identify implementation standards or parameters for their tool that would support regional and national benchmarks. We were unable to identify a common-ground standard for physician attribution; vendors responded positively when we discussed developing standard attribution rules.

Conclusions: It is feasible to identify implementation standards within and across proprietary software solutions measuring cost efficiency or resource use at the physician level for the purposes of benchmarking. Vendors responded positively to participating in the further development of implementation standards or parameters.

Implications for Policy, Delivery, or Practice: With the increasing use of these software solutions by plans and purchasers, and their increasing use for high stakes applications such as for network tiering, vendors and users seem to agree that standardization on these methods would further national policy goals of improving the efficiency of physician directed care.

Primary Funding Source: AHRQ, The Commonwealth Fund

●Qualitative Evaluation of the Implementation of Pay-for-Performance Incentives at Ten Hospitals in Michigan Participating in Rewarding Results

Bert White, D.Min., M.B.A., Bokhour, Barbara, Ph.D, Berlowitz, Dan, M.D., M.P.H., Burgess, James R., Ph.D, Young, Gary, Ph.D, J.D., Sautter, Karen, M.P.H.

Presented By: Bert White, D.Min., M.B.A., Project Director, Health Services Department, BU School of Public Health, 418 High Street, Pembroke, MA 02359; Tel: (781) 826-5674; Fax: (857) 364-4438; Email: bertw@bu.edu

Research Objective: Pay-for-performance financial incentive programs are being implemented across the nation without adequate research to anticipate the resultant interaction dynamics between health plans, payers, and providers. This study evaluated the effect of a health plan-sponsored hospital-based financial incentive program in improving the quality of acute care services by focusing on incentive payouts for one HEDIS quality indicator for the treatment of congestive heart failure. The three research questions were: (1)To what extent do performance-based financial incentives drive quality

improvement efforts in the hospital acute care setting? (2)Are performance-based financial incentives appropriate and effective in improving quality in the provision of care for heart failure in hospital settings? (3)Are there systemic barriers and motivators to the uptake of financial incentives intended to improve quality of care?

Study Design: We conducted hour-long semi-structured interviews with senior managers and cardiologists at ten hospitals participating in the Participating Hospital Agreement (PHA) Program of Blue Cross Blue Shield of Michigan (BCBSM). Under PHA contract, hospitals are eligible for an annual incentive payment of up to 4 percent of BCBSM's DRG-based inpatient claims based on their performance in patient safety, community outreach, and quality of selected indicators. Interview protocols were developed by a multi-disciplinary team of health services researchers, economists, and clinicians using a consensus-building process. Our interview protocol for senior management assessed quality improvement structures and processes, motivators and barriers to participating in PHA, outcomes of participation, and future design on hospital-based incentive programs. In addition, our protocol for cardiologists assessed whether performance in left ventricular ejection fraction assessment was affected by the hospitals participation in the PHA program. We compared national benchmark information, BCBSM-provided data on LVEF assessment rates, and incentive payment amounts over a three-year period with our qualitative findings (2002-2004).

Population Studied: A representative sample of hospitals varying size and performance in LVEF assessment was selected from 86 hospitals participating in PHA. We interviewed 57 senior managers, 9 Chiefs of Cardiology and 1 Director of Cardiovascular Services. All except 1 of the hospitals qualified to receive incentive payment for performance in left ventricular ejection fraction based on PHA thresholds for quality in cardiac care.

Principal Findings: Preliminary findings suggest that financial incentives appear to have an effect in driving quality improvement efforts in acute care settings. Hospital-based programs in larger hospitals in urban settings with higher levels of infrastructure view the incentives as less difficult to earn than smaller hospitals with less supportive infrastructure. In an era of flat or declining reimbursement, senior managers are consistent in their view that financial incentive programs focus hospitals to make investments in systems for quality improvement. For example, Hospital leadership was consistently aware of incentive payments, which ranged from \$40,000 for a small hospital to \$3.7 million for a large medical center. At the same the size of the incentive was not enough to recuperate the costs associated with making quality improvements such as computerized physician order entry systems and taking on additional staff. The majority of cardiologist interviewees reported that 'performance' in LVEF assessment was primarily the result of better documentation rather than improvement in direct patient care. The incentives generated interest in linking office-based patient charts with hospital medical records. Cardiologists are generally unaware of the financial component of the PHA program and reported that they do not receive any of the incentive payment for improved performance. Simultaneously several cardiologists reported direct concurrent benefit from changes in hospital policies including the increase in hospital discharge staff,

chart review nursing, and computerized assistance in guideline adherence

Conclusions: Providing hospitals with financial incentives to improve quality may be an effective approach in motivating hospital administrators to pursue and implement quality improvement programs. There is some evidence to suggest that financial incentives directed at hospitals do not actually change the practice behavior of cardiologists, but lead to improvements in documentation of quality measures.

Implications for Policy, Delivery, or Practice: Hospital-based financial incentive programs, such as PHA, may prove to be effective in motivating hospitals to implement processes and systems for improved quality, which may help hospitals in implementing quality improvement programs that would otherwise be difficult to support. Programs that do not directly incentivize physicians to change their practice behavior for increased fees, but instead, assist physicians in the provision of care may be an appropriate model for future study. Further research on the design and implementation of hospital-based financial incentives to improve adherence to evidence-based guidelines is needed

Primary Funding Source: AHRQ, Robert Wood Johnson Foundation