Revisiting Quality Implications of Efficiency-Based Clinician Profiling

Mark C. Rattray MD; James Andrianos MBA; Dorothy Stam BA

ABSTRACT

BACKGROUND

In our initial study, clinician cost efficiency and process quality compliance appeared to be poor proxies for each other, as measured by leading commercial tools for each domain. One implication was that cost-saving efforts to direct patients to efficient clinicians might have unintended future consequences linked to current deficits in chronic and preventative care. To further understand these findings, we revisited the study, applying improved data processing methods on an expanded data repository.

OBJECTIVE

To improve recent understandings of the relationships between efficiency and quality ratings for individual clinicians, as well as the practical implications that follow.

DESIGN, SETTING, AND MEASURES

Descriptive study of performance measures derived from approximately 88 million claim lines representing a large insurer's entire commercial business during the 3-1/4 year period ending March 31, 2004. Efficiency was defined as condition-specific cost variation over the course of patient care. Quality was defined as process adherence to selected evidence-based medical practice guidelines. Leading commercial tools were employed to assess efficiency and quality performance, respectively. For each type of performance, an index was constructed to summarize an individual clinician's composite achievement.

RESULTS

Correlation between clinicians' efficiency and quality index scores was significant ($p<0.05$) for 5 specialties. For these specialties, efficiency declined as quality compliance increased. Generally, the relationships were weak and do not support inferring one type of performance from the other. For all other specialties, no significant correlations emerged.

CONCLUSIONS

Clinician cost efficiency and process quality compliance again appear to be largely uncorrelated practice characteristics, as measured by the respective and leading commercial tools. In this iteration, we doubled the number of studied claims by enlarging the study period and including both covered and denied services. We emphasized chronic disease and preventative care in an expanded quality analysis. Continuous eligibility requirements were extended to the efficiency analysis. These recurring findings reinforce the recommendation that each practice characteristic should be measured separately.

REVISITING QUALITY IMPLICATIONS OF EFFICIENCY-BASED CLINICIAN PROFILING

MARK C. RATTRAY, M.D.
JAMES ANDRIANOS, M.B.A.
DOROTHY STAM, B.A.‡

High medical cost inflation and stagnant, lackluster quality are fueling creative strategies in the private market for health care. Large self-insured employers are considering "high performance networks" in a bid to moderate medical cost increases. These networks feature clinicians with favorable profiling scores, typically based largely or exclusively on cost efficiency. An implied belief is that efficiency signals quality in health care, but we recently found this relationship to be tenuous at best[1].

In this iteration of the same study, we applied improved data processing methods on an expanded data repository. We focused our quality evaluation on preventative services and chronic disease care, activities likely to lead to unnecessary future costs and complications if overlooked. We again employed the leading commercial measurement tools likely to be employed in private sector interventions.

‡ Mark C. Rattray, M.D., Clinical Faculty, University of Washington School of Medicine; Tel: 206-661-1738; rattray@u.washington.edu.

James Andrianos, M.B.A., Principal, Calculated Risk, Inc.; Tel: 206-382-8575; Fax: 206-382-0581; ja_cri@msn.com.

Dorothy Stam, B.A., Manager, Clinical Performance Measurement, Regence BlueShield; Tel: 206-287-5973; Fax: 206-287-5844; dtstam@regence.com.

The authors recognize Zakaria Ramadan-Jradi, M.D., M.B.A., for his significant guidance and insight into clinical performance measurement.

METHOD

We examined measurable relationships between the quality and efficiency of individual clinicians. We defined quality as rates of process adherence to selected evidence-based medical practice guidelines. We examined cost in terms of efficiency, which was defined as condition-specific cost variation over the course of patient care. We relied on a single large repository of detailed claims to study in parallel the quality and efficiency performance of discrete clinicians. Using the claim repository, quality care was identified and credited to individual clinicians. The repository also permitted each patient's total medical cost to be parsed into its constituent clinical episodes of care and their corresponding costs. Clinicians were compared only to other clinicians of the same specialty.

Data Collection. We obtained 39 months of detailed claim records comprising the entire commercial "book of business" for Regence BlueShield, an independent licensee of the Blue Cross and Blue Shield Association in Washington State. The records represented the care delivered to approximately 1.4 million patients and amounted to nearly 88 million claim lines over a 3-1/4 year period ending March 31, 2004. Claims spanned both inpatient and outpatient care, whether delivered via professional, facility, or ancillary providers. All pharmacy claims were included.

Quality Evaluation Tool and Analysis. We utilized the services of Health Benchmarks, Inc. to conduct algorithmic examinations of the claim repository. The algorithms measured process compliance with evidence-based medical guidelines. Algorithms for each indicator were adapted from existing published and peer-reviewed protocols comprising modern evidence-based medical guidelines.

Exhibit 1 lists the 42 process indicators for which compliance was evaluated. For each indicator, a rate of compliance was determined for every qualifying clinician within the applicable specialty. This raw compliance rate was converted to a percentile score relative to the performance of other clinicians with the same specialty. Across indicators, all of a clinician's percentile scores were averaged into a single composite score. A clinician's quality index was his/her composite score relative to the median. An index value of 1.0 approximated the statewide network median performance within a specialty, and therefore, not necessarily a performance target.

Efficiency Evaluation Tool and Analysis. Using the same claims, distinct clinical episodes of care were identified using dedicated grouping software. We licensed a clinical episode grouper from Symmetry Health Data Systems, Inc.. The grouper examines individual claim lines for clinically-related services, combining them into episodes of varying duration, depending on the appearance of pre-defined "quiet periods" in the claim sequence. The resulting episodes may be isolated, sequential, or concurrent. The grouper maintains diagnostic profiles of each patient, accounting for co-morbid conditions during Episode Treatment Group® (ETG) assignment. Each ETG approximates a clinically homogeneous classification, permitting case mix adjustment when comparing episode sets of varying ETGs. There are 850 distinct ETGs.

We derived relative efficiency scores for individual clinicians. For each episode, we identified the clinician most responsible for professional care, and we attributed the episode to this clinician. Clinicians with sufficient volumes of attributed episodes became eligible for efficiency comparison. For a given clinician's attributed episodes, expected resource utilization was compared to observed resource utilization, as measured by paid dollars. Paid dollars included insurer reimbursement, patient deductibles, and co-insurance. The efficiency index was the ratio of expected-to-observed episode costs. An index value of 1.0 approximated the statewide network average within a specialty, and therefore, not necessarily a performance target.

Detailed documentation of the methods used to collect, prepare, process, and analyze the claims in this study is available in the attached Technical Appendix.

RESULTS

We derived both quality and efficiency index scores for 3,756 clinicians representing 26 specialties. We examined the relationship between quality and efficiency index scores for each clinician by measuring the significance of within-specialty linear models. Exhibit 2 presents results from applying linear regression analysis to clinicians' quality and efficiency index scores by specialty.

Of 26 studied specialties, 5 models exhibited statistically significant ($p<0.05$) linear fits. Fourteen specialties had statistically insignificant relationships.

We categorized an additional 7 specialties as lacking sufficient data, using a minimum of 25 clinician observations as a threshold. The 19 specialties with sufficient data were:

- Adult Medicine (nurse practitioner)
- Allergy
- Cardiovascular Disease
- Emergency Medicine
- Endocrinology
- Family Practice
- Family Practice (nurse practitioner)
- Family Practice (physician assistant)
- General Practice
- Internal Medicine

- Midwife/Obstetrics/Gynecology (nurse practitioner)
- Naturopath
- Obstetrics/Gynecology
- Oncologist/Hematologist
- Ophthalmology
- Pulmonary Diseases
- Pediatrics
- Urology
- Women's Health Practitioner (nurse practitioner)

Residuals for these specialties showed no evidence of non-linearity.

The 5 specialties with significant relationships were Emergency Medicine, Endocrinology, Internal Medicine, Obstetrics/Gynecology, and Ophthalmology. Their results suggest quality improves as efficiency declines. Although relationships were significant, efficiency and quality appear to be poor predictors of one another (Exhibit 3).

COMMENT

In this iteration of this study, we again conclude that clinician-level process adherence to evidence-based medical guidelines is largely uncorrelated with episode-of-care cost efficiency, as measured by widely used commercial tools and methods. Weak exceptions exist in certain specialties, where quality gains tend to correspond with declining efficiency, a directional relationship also seen in the previous study.

Limitations and Opportunities. Our methods do not adjust for efficient but inappropriate treatment decisions. Unlike compliance with clinical process guidelines, the existence of identified episodes of care need not validate the appropriateness of chosen treatment options.

The earliest process indicators to be converted into claim algorithms were part of the Health Plan Employer Data and Information Set (HEDIS®). A portion of the quality indicators used in this study is HEDIS® or HEDIS®-like in nature. Most of these indicators apply to primary care. We reiterate our call for the development of quality indicators applicable to surgical and procedural specialties.

Without recognized and open standards for measuring professional quality and efficiency, we opted to employ the tools of leading commercial vendors, complimented by certain methodologies we have attempted to describe here. We applaud and encourage multi-stakeholder efforts to build consensus around measurement standards and to enhance methods in this area, such as the White Paper recently sponsored by The Leapfrog Group, Bridges to Excellence, and The Commonwealth Fund[2].

Implications. Measuring quality and efficiency separately with dedicated tools remains the prudent approach to gauging these aspects of clinician performance.

Funding. The Regence Group, an independent licensee of the Blue Cross and Blue Shield Association funded the design and conduct of this study; the collection, management, analysis, and interpretation of the data; and the preparation, review, and approval of the manuscript.

NOTES

1. See 2004 version of this study, *Quality Implications of Efficiency-Based Profiling* at www.regence.com/research.
2. See *Measuring Provider Efficiency (Version 1.0): A Collaborative Multi-Stakeholder Effort* at www.regence.com/research.

Exhibit 1: Selected Indicators of Conformance with Evidence-based Medicine	*Source: Health Benchmarks, Inc.*
Process Indicator	Number of Applicable Specialties
Diabetic retinal exam	24
Subsequent Annual Screening for Diabetic Retinopathy	24
Antidepressant Medication Management: Acute Phase Management	23
Glycosylated hemoglobin (HbA1c) for diabetics	23
Lipid panel for diabetes	23
Screening for diabetic nephropathy	23
Use of long-term asthma control drugs	23
Optimal Treatment for Asthmatics with High Use of Rescue Medications	23
Treatment/Follow-up after emergency department visit for asthma	23
X-ray prior to knee MRI	23
ER utilization of asthma patients	23
Inappropriate Use of Antiobiotics for Viral URI	22
First line antibiotics for acute otitis media	22
Follow-up after hospitalization for mental illness	22
Antidepressant Medication Management: Optimal Practitioner Contacts	22
First line antibiotics for acute sinusitis	22
First line antibiotics for tonsillitis	22
Cervical cancer screening	22
X-ray prior to MRI or CAT scan of lower back pain	22
Compliance with Inhaled Corticosteroids for Moderate to Severe Asthmatics	22
Antidepressant Medication Management: Continuation Phase Management	22
Pregnancy test on patients with vaginal bleed or abdominal pain	21
Short-course antibiotic therapy for uncomplicated UTI in women	21
Treatment of coronary artery disease: lipid lowering drugs	20
Follow-up after initial diagnosis and treatment of colorectal cancer	20
Chlamydia screening for women	20
Treatment of coronary artery disease: ACE/ARB	20
Treatment of coronary artery disease: beta blockers	20
Avoidance of First Line Use of Calcium Channel Blockers for CHF	19
ACE inhibitor use in congestive heart failure	19
Regular monitoring of electrolytes and renal function for patients taking digoxin	19
Compliance with lipid-lowering drugs	19
Duretics/beta blockers as 1st line therapy: new antihyptsn pharmacotherapy starts	19
Hepatic enzyme monitoring for statin use	19
Compliance with anti-hypertensive drugs	19
Visual Field Test for Patients with Suspected Glaucoma	17
Lipid panel following stroke	17
Compliance with glaucoma medications	17
Annual CBC or hemoglobin/hematocrit for pre-ESRD patients	17
Annual creatinine & potassium tests for pre-ESRD patients	17
Mammography screening	17
Annual visual field tests for glaucoma patients	17

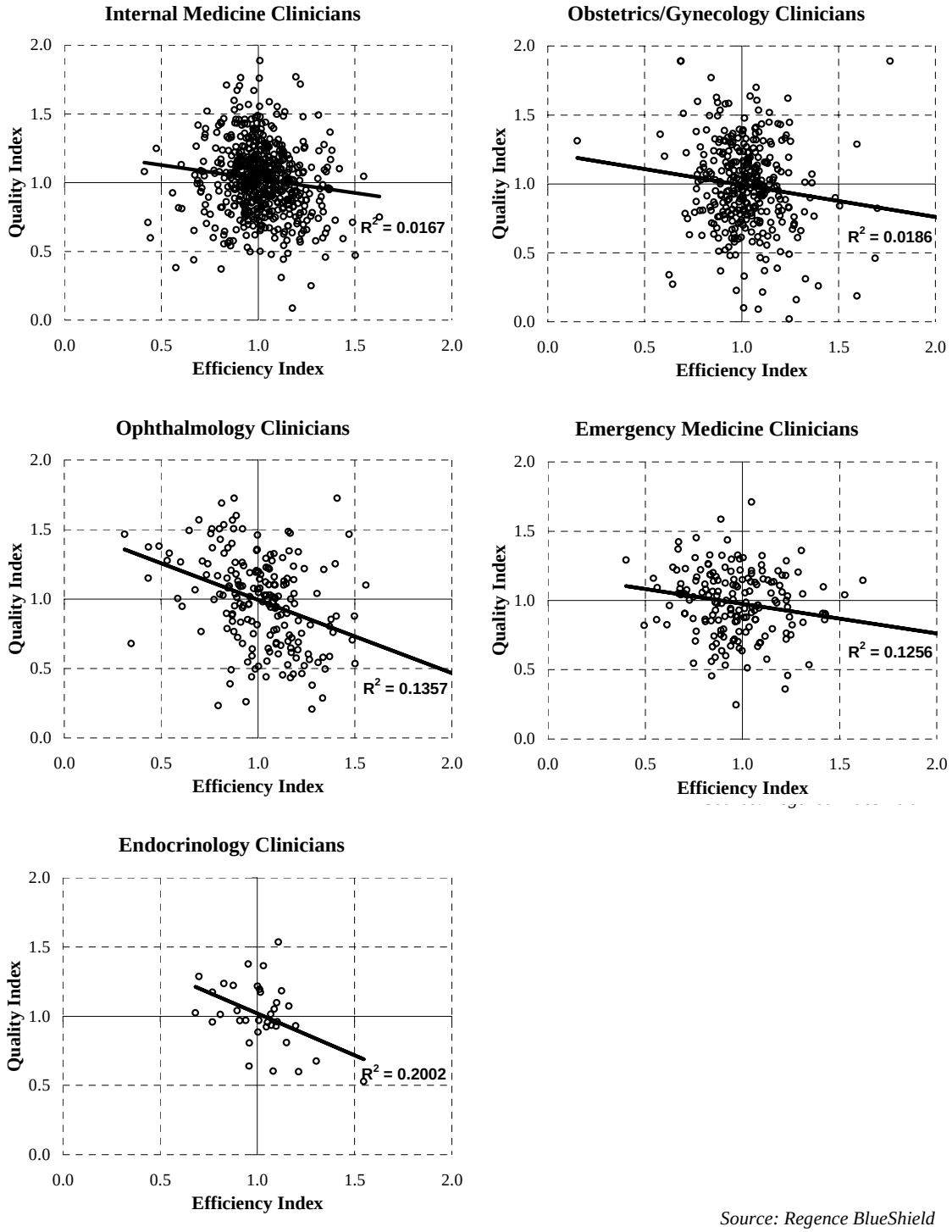
Exhibit 2: Linear Regression Results of Clinician Quality and Efficiency Index Scores

Specialty	Clinician Count	Applicable Quality Indicators	p-value	Coefficient	Lower 95% C.I.	Upper 95% C.I.	R^2
Significant (p<0.05)							
Internal Medicine	564	42	0.002	(0.203)	(0.333)	(0.074)	0.0167
Obstetrics/Gynecology	341	41	0.012	(0.232)	(0.412)	(0.052)	0.0186
Ophthalmology	191	6	<.0001	(0.526)	(0.717)	(0.336)	0.1357
Emergency Medicine	181	20	<.0001	(0.214)	(0.298)	(0.131)	0.1256
Endocrinology	36	13	0.006	(0.604)	(1.024)	(0.183)	0.2002
Not significant (p>=0.05)							
Family Practice	1,215	42	0.210	(0.067)	(0.172)	0.038	0.0013
Pediatrics	260	20	0.487	0.199	(0.365)	0.763	0.0019
Family Practice (N.P.)	180	42	0.063	(0.221)	(0.455)	0.013	0.0192
Family Practice (P.A.)	157	41	0.385	(0.087)	(0.284)	0.110	0.0049
Cardiovascular Disease	123	16	0.833	0.009	(0.077)	0.096	0.0004
General Practice	107	42	0.106	(0.249)	(0.551)	0.053	0.0247
Naturopath	77	41	0.903	0.021	(0.322)	0.364	0.0002
Midwife/Ob/Gyn (N.P.)	38	5	0.926	0.041	(0.852)	0.934	0.0002
Urology	38	4	0.388	(0.374)	(1.243)	0.495	0.0208
Oncologist/Hematologist	37	2	0.378	(0.090)	(0.294)	0.114	0.0223
Allergy	34	9	0.871	(0.039)	(0.523)	0.445	0.0008
Women's Health (N.P.)	33	41	0.516	0.285	(0.599)	1.169	0.0138
Pulmonary Diseases	28	6	0.764	(0.071)	(0.552)	0.410	0.0035
Adult Medicine (N.P.)	27	42	0.102	(0.885)	(1.959)	0.189	0.1033
Insufficient data (less than 25 observations)							
Osteopath	23	41	0.868	(0.034)	(0.456)	0.387	0.0014
Gynecology	16	5	0.518	(0.413)	(1.748)	0.922	0.0305
Vascular Surgery	16	5	0.547	0.129	(0.319)	0.577	0.0264
Ob-Gyn (P.A.)	10	41	0.452	(1.238)	(4.850)	2.373	0.0725
Otorhinolaryngology	9	4	0.030	0.693	0.091	1.296	0.5142
Pediatrics (N.P.)	8	18	0.191	2.779	(1.837)	7.395	0.2656
Psychiatry	7	4	0.064	(1.560)	(3.250)	0.130	0.5297
Total	**3,756**						

N.P. = Nurse Practitioner P.A. = Physician Assistant

Source: Regence BlueShield

Exhibit 3: Scatter Plots of Specialties with Significant (p<0.05) Quality-Efficiency Relationships
Note: index scores greater than 1.0 indicate above average quality or efficiency



Source: Regence BlueShield

TECHNICAL APPENDIX

Figure 1 provides a Data Processing Overview, referenced throughout this section.

Data Collection. We extracted 39 months of claims for the commercially-insured, largely Preferred Provider and Point of Service populations of Regence BlueShield, an independent licensee of the Blue Cross and Blue Shield Association in Washington State. Approximately 1.4 million patients generated nearly 88 million claim lines over a 3-1/4 year service period through March 31, 2004. Claims spanned both inpatient and outpatient care, whether delivered via professional, facility, or ancillary providers. All pharmacy claims were included. Medicare, Medicaid, and government program claims were excluded, as were capitated programs. Each claim line included data elements found in typical adjudication systems, such as patient identifiers, clinician identifiers, service dates, eligibility dates, procedure codes, and diagnosis codes.

Denied claims were included. Analyzing denied claims enhances the efficiency evaluation by extending the duration of a single episode that might be identified incorrectly as multiple, sequential episodes. The quality analysis potentially benefits from a more accurate accounting of compliance with standards of recommended care.

The same claims were used in evaluating process quality and cost efficiency.

PROCESS QUALITY

Quality Data Preparation. Health Benchmarks measured the frequency of distinct claim values in the procedure code, diagnosis code, and National Drug Code data fields. Frequencies were compared with expectations from other commercial plans' claims. Where significant deviations appeared, including errant, unknown, or proprietary codes, an investigation was conducted. Sometimes, investigations resulted in revisions to the selected process indicators.

Eligibility history tables were evaluated to confirm that continuous patient eligibility reflected expected levels (Fig. 1, Cell Q1).

Quality Data Processing. We selected 42 process indicators for evaluation. Selection criteria included overall coverage across the professional specialties, and clinical relevance given the commercial insurance-aged population. Some indicators apply to a single specialty; others apply to several. For instance, X-ray Prior to Knee MRI applies to General Practice, Family Medicine, Emergency Medicine, and Orthopedic Surgery. The number of indicators applicable to a single specialty ranged between 1 and 41.

Patient-specific enrollment studies were conducted to assure that cases used in quality evaluation featured patients who were sufficiently continuously eligible for benefits. Where necessary, each indicator's algorithm was adapted to account for idiosyncratic claim aspects, such as proprietary procedure codes. The algorithms for all indicators were then run against the entire claim repository. Output for each indicator was examined for unexpected results, such as unusual implied disease prevalence rates. Any anomalies were investigated, algorithms refined, and rerun (Fig. 1, Cell Q2).

Assignment of cases to clinicians utilized a patient-centered, team-based approach. For each indicator, a case was opened for a clinician when:

1. The indicator in question applied to the clinician's specialty

2. The clinician had an office visit with the patient in question
3. The patient was continuously eligible for benefits

As an example, consider the mammography indicator. Suppose a continuously eligible patient has office visits to both a family practitioner and a gynecologist. A case is opened and assigned to both clinicians. If one or both clinicians meet the mammography recommendation, both clinicians receive compliance credit (Fig. 1, Cell Q3).

Quality Data Analysis. A raw compliance rate was calculated for each clinician-indicator combination. Clinicians meeting the following requirement were evaluated for quality:

- Each specialty-indicator combination must have contained at least 100 cases network-wide
- Only clinician-indicator combinations with at least 10 cases were retained
- Only indicators featuring at least 10 different clinicians of the same specialty were retained
- Of retained cases, a clinician must have had at least 30 cases across all indicators to be evaluated.

The clinician's raw compliance rate was restated as a percentile score, reflecting the performance of same-specialty clinicians for the given indicator. Each clinician's percentile scores from multiple indicators were averaged into a single composite score. The composite score was indexed such that 1.0 approximated the median network compliance as shown in Example 1.

Example 1: Quality Index Calculation

Requirements:
(1) Each Specialty-Indicator combination must contain at least 100 cases network-wide
(2) Only clinician-Indicator combinations with at least 10 cases are retained
(3) Only Indicators featuring at least 10 different clinicians of the same specialty are retained
(4) Of retained cases, a clinician must have at least 30 cases across all Indicators to receive a composite score

Doctor 001, General Practice	A	B	C	D	E
		CLINICIAN		**GEN. PRACTICE**	
Health Benchmarks, Inc. Quality Indicator	**Cases**	**Success Rate**	**Peer Percentile**	**Median Percentile**	**Quality Index (C) / (D)**
Compliance with lipid-lowering drugs	10	-	1%	50%	0.02
Screening for diabetic nephropathy	18	0.17	35%	50%	0.70
Inappropriate Use of Antibiotics for Viral URI	49	0.31	15%	50%	0.30
Mammography screening	11	0.36	1%	50%	0.02
Compliance with anti-hypertensive drugs	25	0.48	3%	50%	0.06
Glycosylated hemoglobin (HbA1c) for diabetics	18	0.50	66%	50%	1.32
Diabetic retinal exam	18	0.56	91%	50%	1.82
Cervical cancer screening	16	0.63	12%	50%	0.24
Lipid panel for diabetes	18	0.67	31%	50%	0.62
Overall:	**183**	**0.41**	**28%**	**50%**	**0.57**

In this way, we derived an aggregate, relative performance index for each clinician. There were 4,658 clinicians (31 specialties) with quality index scores (Fig. 1, Cell Q4).

Only within-specialty clinician comparisons are relevant. Comparing the quality scores of clinicians with different specialties can be problematic because the applicable process indicators vary between specialties.

Comparisons were made to the median statewide performance. This baseline is not intended to represent a standard of practice.

COST EFFICIENCY

Cost Efficiency Data Preparation. The episode grouping software requires each claim line to be assigned to 1 of 3 Provider Types: Facility, Clinician, or Other. Facility-type providers include hospitals, skilled nursing facilities, etc. The Clinician category encompasses professionals who diagnose, treat, and recommend patient treatment. The Other category includes remaining providers: laboratories, pharmacies, ambulance, etc.

Each claim line must also be assigned to 1 of 3 Types of Service: Room and Board, Medical-Surgical, or Ancillary. Room and Board generally designates inpatient confinements. Medical-Surgical normally denotes clinician-delivered services. Ancillary contains all remaining services.

The grouping software in part relies on Provider Type and Type of Service assignments to identify encounters justifiably beginning or extending a clinical episode. We prepared the claim extract with Provider Type and Type of Service assignments. (Fig. 1, Cell E1).

Cost Efficiency Data Processing. Once the claim extract was prepared, episode grouping occurred. The grouper examined individual claim lines for related clinical encounters. Within collections of clinically related encounters, the software examined the claim sequence for periods without service activity. These "clean periods" vary by ETG and determine whether subsequent clinical activity should be considered part of an existing episode or the start of a new episode.

Only complete episodes were recognized during the grouping process. Episodes already underway before the earliest service date in the claims extract were not recognized; episodes still open by the latest service date were rejected. An exception was made to capture chronic conditions. Chronic episodes exceeding 1 year were split into sequential episodes lasting up to 1 year.

Thus, the ETG software grouped clinically related encounters within the claim extract. Grouping results were checked against benchmarks and found to be within expected ranges. Benchmarks included frequency of ungrouped claim lines; distribution of complete and incomplete episodes; frequency of errant diagnosis, procedure, and drug codes; etc. (Fig. 1, Cell E2).

An important step in measuring efficiency is defining an episode attribution rule: the method for assigning an entire episode to the clinician most responsible for its constituent services, whether rendered, ordered, or referred to others. We used a modified version of a widely used attribution rule, which assigns the episode to the clinician with the greatest share of charges. We made 1 modification and 2 qualifications:

1. Use of paid dollars, rather than charges, thereby capturing total expected clinician reimbursement, including patient cost sharing
2. Restriction of the examined services to professional records only (office visits and medical/surgical procedures)
3. Requirement that an attributed clinician account for at least 30% of the total paid dollars for these examined services.[1]

For example, consider a complete episode comprising $1,000. Suppose professional office visits and surgical procedures amount to $450. If no clinician accounted for at least $135 of this service subset, or 30%, no episode attribution occurs. The clinician with the highest percentage greater than or equal to 30% is assigned the episode. When 2 or more clinicians are tied for the greatest percentage, no attribution occurs.

The episode grouping software does not refer to patient-specific eligibility patterns when identifying episodes of care. Rather, it describes the episode's start and end based on the appearance of pre-defined "quiet periods" in the claim sequence. Because the absence of claim activity might be created by a period of ineligibility, we removed episodes that did not meet the following continuous eligibility standards:

1. For each episode, we determined the patient's maximum continuous break in eligibility during the period covered by the episode, plus its designated beginning and ending quiet periods. When the period of maximum continuous ineligibility exceeded the length of the quiet period, we rejected the episode. This standard reduces the possibility that a break in eligibility will be mistaken for a quiet period.
2. For each episode, we calculated the ratio of ineligible days to total days in the period covered by the episode, plus its designated beginning and ending quiet periods. When the ratio exceeded 12.3%, we rejected the episode. This threshold originates in a commonly used rule for assessing eligibility continuousness that permits up to 45 ineligible days per year.

Data loss during claim submission and processing can result in aberrant episodes with very low costs. Within each ETG, we identified and eliminated these "low outlier" episodes from further analysis.

Because a clinician's specialty establishes his/her peer group, it is important to confirm that *declared* specialties do not differ from *effective* specialties. We examined the nature of episodes attributed to primary care clinicians to evaluate differences between declared and effective practicing specialty. We employed a higher-level summary of ETGs called MPCs (Major Practice Category). MPCs roughly correspond to professional specialty designations. When more than half of a primary care clinician's attributed episodes occurred in a single, non-primary care MPC, we reassigned the clinician to that specialty's peer group.

Unusually high cost episodes can reflect extraordinary patient-specific circumstances. We examined and eliminated a fixed fraction of high cost episodes for each combination of clinician and ETG to reduce this source of variation (Fig. 1, Cell E3).

We used version 5.0 of the Symmetry episode grouper. Data preparation and analysis activities were conducted using SAS for Windows version 8.2, with SAS Enterprise Guide version 2.0 as a graphical user interface. The computing platform was a dedicated Hewlett Packard Proliant DL 380 G3 with Intel Xeon 3.06 GHz processing and 4 gigabytes of RAM.

Cost Efficiency Data Analysis. Table A provides an overview of the data exclusion steps described in this section.

Table A: Data Exclusion Overview

Data Stage	Paid Dollars		Episodes		Unique Clinicians	
All Studied Claims	**$6,890,196,880**		**9,536,968**		**32,774**	
Claims Grouped to Complete Episodes	$4,362,781,901	100%	7,506,425	100%	28,848	100%
Episodes with Clinician Attribution	$3,135,318,926	72%	4,756,499	63%	27,391	95%
Patient is Continuously Eligible	$2,122,177,650	49%	3,741,168	50%	24,891	86%
Episode is not a Low Outlier	$2,109,830,269	48%	3,648,319	49%	24,684	86%
Episode is not a High Outlier	$1,321,632,351	30%	2,986,577	40%	15,746	55%
>=100 Episodes Exist for each ETG-Specialty	$1,171,027,226	27%	2,881,858	38%	14,668	51%
>=10 Episodes Exist for each Clinician-ETG	$858,382,460	20%	2,305,488	31%	8,653	30%
>=10 Clinicians Exists for each ETG-Specialty	$782,136,269	18%	2,233,182	30%	8,397	29%
>=30 Episodes Exist for each Clinician	$757,762,607	17%	2,201,618	29%	6,547	23%

From 87,613,641 claim lines, 7,506,425 distinct, complete episodes were identified, representing 53,272,264 claim lines (or about 61%). Of these 7,506,425 episodes, we eliminated 4,624,567 episodes in the following categories:

- Attribution rule resulted in no clinician assignment (2,749,926)
- Patient did not meet continuous eligibility requirements (1,015,331)
- Low outliers (92,849)
- High outliers (661,742)
- ETG-Specialty combinations with less than 100 episodes (104,719)

The remaining 2,881,858 episodes constituted the basis for deriving clinician efficiency index scores. These episodes accounted for $1,171,027,226, or 27% of the all paid dollars in the episode-grouped claims.

For these 2,881,858 episodes, the network-wide average paid dollars per episode was calculated for each combination of specialty and ETG, representing the expected cost. Clinicians meeting the following requirement were evaluated for efficiency:

- Only clinician-ETG combinations with at least 10 attributed episodes were retained
- Only ETGs featuring at least 10 different attributed clinicians of the same specialty were retained
- Of retained episodes, a clinician must have had at least 30 attributed episodes of any ETG to be evaluated.

These clinicians' collective episode volume was 2,201,618, accounting for $757,762,607.

To measure a specific clinician's efficiency, the network expected cost for each attributed episode was compared to the clinician's corresponding observed cost. The ratio of total expected to total observed dollars became the clinician's relative efficiency index as demonstrated in Example 2 below. Note that an index score of 1.0 approximated the network-wide expectation for the studied mix of ETGs.

Example 2: Efficiency Index Calculation

Requirements:
(1) Each Specialty-ETG combination must contain at least 100 comparative episodes network-wide
(2) Only clinician-ETG combinations with at least 10 attributed episodes are retained
(3) Only ETGs featuring at least 10 different attributed clinicians of the same specialty are retained
(4) Of retained episodes, a clinician must have at least 30 attributed episodes across all ETGs to receive a composite score

Doctor ABC, Family Practice	A	B	C	D	E: D / C	F: E x A
Episode Treatment Group®	**Clinician Episodes**	**Clinician Cost**	**Network Episodes**	**Network Cost**	**Average Network Cost**	**Expected Cost: Average Network Cost * Clinician Episodes**
Otitis media, w/o surgery	14	$967	25,679	$2,429,351	$95	$1,324
Tonsillitis, adenoiditis or pharyngitis, w/o surgery	19	$1,707	43,417	$3,957,743	$91	$1,732
Acute bronchitis, w/o comorbidity, age 5+	17	$2,250	29,506	$3,708,763	$126	$2,137
Total:	**50**	**$4,924**				**$5,193**
Efficiency Index Score (Sum F / Sum B):						**1.05**

In this way, we derived an aggregate, relative performance index for each clinician. There were 6,547 clinicians (49 specialties) with efficiency index scores (Fig. 1, Cell E4).

We opted not to compare the treatment patterns of physicians with differing specialties, even when treating the same condition. For example, we recognized an episode of sinusitis attributed to an otolaryngologist separately from one attributed to a family medicine physician. This acknowledgement enhances within-specialty severity adjustment.

As with quality comparisons, efficiency comparisons were made to statewide performance. This baseline is not intended to represent a standard of performance.

There were 4,658 clinicians (31 specialties) with quality index scores and 6,547 clinicians (49 specialties) with efficiency index scores; 3,756 clinicians (26 specialties) had both quality and efficiency index scores. A total of 32,774 distinct clinicians (173 specialties) comprised the professional network during the study period. Thus, we evaluated 11.5% of the total clinician network for both process quality and cost efficiency.

NOTES

1. Past analysis showed a low frequency of episodes with multiple managing clinicians, regardless of the attribution threshold under consideration. At attribution thresholds beyond the 35%-40% range, we begin to see an increasing trend where the more costly episodes are not attributed to a responsible clinician. See *Episode Attribution Study* at www.regence.com/research.

Figure 1 (Technical Appendix): Data Processing Overview for Deriving Professional Quality and Cost Efficiency Index Scores

DATA COLLECTION

Selected Claims:
3-¼ year commercial [Preferred Provider and Point of Service], non-Medicare, non-Medicaid, book-of-business claim line detail. Includes professional, facility, ancillary, and drug claims. Claims lines total approximately 88 million.

Obtain Twin Copies of Detailed Claim Extract

PROCESS QUALITY | COST EFFICIENCY

DATA PREPARATION

Preparation Activities:
Compare frequencies of specific values in claim record with other insurers' books-of-business. Test rates of continuous member eligibility with benchmarked expectations.

Q1
Claim Examination and Preparation for Quality Analysis

E1
Claim Examination and Preparation for Cost Efficiency Analysis

Preparation Activities:
Assign each claim line to standard Provider Type and Service Type values.

DATA PROCESSING

Patient-Level Quality Evaluation :
Adapt claim search algorithms for insurer-specific issues. Run algorithms against claims and examine output. Investigate anomalies, refine algorithms, and rerun as necessary. Confirm continuous eligibility of studied patients.

Patient-Clinician Case Assignment:
A case is opened for each office visit. Each indicator could apply to multiple clinicians treating a patient. All clinicians of appropriate specialty receive credit if the process standard is met.

Q2
Evaluate Cases for Process Quality

E2
Identify Distinct Clinical Episodes

Episode Grouping Process:
Identify complete episodes only. Confirm that records grouping to episodes (un)successfully meet benchmarked expectations.

Episode Attribution to Clinicians:
Attribute episode to clinician with greatest share of professional payments larger than 30%, or discard episode. Confirm continuous eligibility of studied patients. Remove low outlier episodes. Test that "declared" and "effective" specialties match. Remove high outlier episodes.

Q3
Identify Cases and Assign to Clinicians

E3
Attribute Episodes to Clinicians and Remove Problematic Episodes

DATA ANALYSIS

Quality Index Score:
Requirements:
(1) Each specialty-indicator combination must contain at least 100 cases network-wide
(2) Only clinician-indicator combinations with at least 10 cases are retained
(3) Only indicators featuring at least 10 different clinicians of the same specialty are retained, and
(4) Of retained cases, a clinician must have at least 30 cases across all indicators to be evaluated

The clinician's raw compliance rate is restated as a percentile score relative to same-specialty clinicians. Each clinician's percentile scores from multiple indicators are averaged into a single composite score. The composite score is indexed such that 1.0 represents the median network compliance.

Q4
Calculate Clinicians' Individual Quality Index Scores

E4
Calculate Clinicians' Individual Cost Efficiency Index Scores

Examine Correlation Between Process Quality and Cost Efficiency Scores

Statistical Examination:
Examine within-specialty linear regression models for clinicians with both quality and cost efficiency index scores. Test for coefficient significance better than p=0.05. Study residuals for evidence of non-linearity.

Efficiency Index Score:
Requirements:
(1) Each specialty-ETG combination must contain at least 100 comparative episodes network-wide
(2) Only clinician-ETG combinations with at least 10 attributed episodes are retained
(3) Only ETGs featuring at least 10 different attributed clinicians of the same specialty are retained, and
(4) Of retained episodes, a clinician must have at least 30 attributed episodes of any ETG to be evaluated

The ratio of total expected to total observed dollars is the clinician's relative cost efficiency index. An index score of 1.0 approximates the network-wide performance for the studied mix of ETGs. Index scores higher than 1.0 indicate above average cost efficiency.