Issue Brief

Piercing The Darkness: A Generalizable Approach To Reliably Measuring Quality Of Care

Francois de Brantes, Andrew Wilson, Amita Rastogi, MD

Background

In any other market, services and products that are supplied to consumers are well advertised and evaluated. Before purchasing any expensive item, most individuals tend to do their research, become informed and educate themselves on their options. Consumers are eager to know if they are getting the best product they can for the best price possible.

The healthcare market should function in the same manner, and yet it can't because there is a significant lack of consistently available physician-level quality of care data.[1, 2] As a result, consumers can't really assess the quality of care they are receiving, nor are there easy ways of conducting research to make an informed decision before deciding to which hospital or doctor to entrust their wellbeing.[3] To an extent, the healthcare market is akin to a consumer entering a store blindfolded. They have no means of knowing if they are choosing the right item, if they are getting what they need or what price that item is even worth.

For all intents and purposes, the healthcare market is an economic failure. In the words of an economist: "The features of markets described in economic textbooks[4] are not found in the healthcare industry and thus inhibit efficient operations of supply and demand. These features include lack of price information and pricing transparency; lack of data on product quality; the resulting inability to assess the comparative value (defined as quality divided by cost) of products and

services; asymmetric information between providers and consumers; imperfect agency relationships between physicians and their patients; the heavy role of government as both a buyer and regulator; and moral hazard flowing from insurance coverage leading to distortions in market efficiency."[5, 6] Much of these distortions can, in fact, be straightened out with transparency, because better informed individuals lead to better health outcomes and lower healthcare costs.[7] And if the benefits are so clear, why is there a persistent lack of transparency?

For years the process of defining measures for public use has been dominated by physician and hospital organizations as well as the federal government through its mandated reporting programs. The politicization of the process has generated a parsimonious set of quality measures that, by and large, fail to differentiate provider performance.[8] Moreover, there are significant gaps in what is being measured and reported with respect to relevance to under 65, commercially insured Americans. A recent report by Catalyst for Payment Reform highlights the deficiencies of the current measurement system and proposes some priority measures for employers and other purchasers to act on.[9,10] The upshot is that we lack quality measures on physicians, at a level that makes sense to the average consumer, because the national physician leadership organizations have resisted comprehensive and differentiating measurement. And unfortunately, payers in the public and private sectors, have failed to push back against "organized medicine" and demand greater accountability for the quality of care delivered in the US.

In response to latent consumer demand for physician quality measures and the unwillingness of policy-makers and industry leaders to respond to that demand, some not-for-profit organizations have stepped up. In our annual Physician Quality Transparency Report Card[8] we highlight community-based organizations that are collecting and reporting physician quality. In addition, ProPublica[11] and Consumers Checkbook[12] have both released quality ratings of surgeons on a defined set of hospital-based surgeries. And since the data used to calculate these ratings comes from Medicare, it is mostly focused on surgeries that are prevalent in that population and therefore not necessarily relevant to all Americans.[13]And yet, it's the best the country has to offer to consumers seeking comparative information on the quality of physicians.[14]

With the advent of all-payer claims databases and statewide efforts to reform payment and delivery systems, the opportunity to push for greater physician quality transparency is at hand. The charge for all of us is to find a way to get it done in a manner that will be acceptable to consumers while keeping providers engaged. The first step is to understand how to best frame quality measures for consumers, and to understand the features of quality measurement and reporting that are important to consider when publishing ratings.

Framing Quality Measures

Organizations like Consumer Reports have long understood through careful research how to frame quality metrics for all types of goods and services. The iconic half circles and full circles are the result of years of consumer testing on attitudes towards how to present quality data in an understandable way. Rating consumer goods, however, is not really dependent on the acceptability of the measures by the manufacturers of the products being measured. It was Ralph Nader and not the automobile industry who first called public attention to data that evidenced poor safety of American autos. The industry itself ultimately embraced the notion that safety matters, and went on to develop additional measures – but it was an external actor that got things started. Rating physicians or hospitals is, however, a different enterprise. That's because patients interact with physicians and hospitals in a very different way than most other service providers.

The recent backlash to ProPublica's surgeon ratings can serve as an example for the pitfalls of measuring individual physicians.[15] Starting in late 2014, HCI[3] convened a panel of experts in quality measurement as well as experts in different domains of measurement, including patient attitudes to quality data. The group was tasked with identifying the critical features of quality measures that would make them acceptable to consumers and physicians. While the list is very similar for both, there are differences in how important each feature is to each group. The features, by order of importance, are summarized in Table 1.

TABLE 1: Essential Features of Quality Measures by Stakeholder

IMPORTANT FEATURES FOR CONSUMERS:	IMPORTANT FEATURES FOR PHYSICIANS:
1. Measuring outcomes of care	1. Understanding who's being measured
2. Distinguishing performance between providers	2. Having benchmarks
3. Having benchmarks	3. Measuring outcomes of care
4. Understanding who's being measured	4. Ensuring appropriate risk adjustment
5. Clearly stating what condition, procedure or population is being measured	5. Using fully transparent methods for measuring, adjusting and rating

Overall, what's important to consumers is intuitively understandable. They want to know whether they're getting good care or bad care. Today's commonly used "process measures" which simply tell consumers whether a test or screening was performed is simply not useful. They also want to clearly distinguish hospitals or physicians with simple rating systems like Consumer Reports or stars, and have benchmarks to understand whether the providers in their area are, overall, better or worse than national averages (or the national top quartile). And finally they want

to know who exactly is being measured, whether it's a facility, a practice, or an individual clinician, and what's being measured, whether it's an entire population, a single condition, or a procedure. Their preference is for facility measurement for the medical events that are staged in hospitals, and physicians for everything else. That, of course, contrasts and conflicts somewhat with what's important to physicians.

Overall, physicians prefer that measurement be at the practice, medical group, hospital or health system level. They don't particularly like individual physician measurement. They want benchmarks of performance so that they can compare theirs, in absolute terms, to the benchmark, because that's more important to them than simply knowing they're average, or below/above average. While they have preferred that public reports be on process measures, most physicians agree that outcome measures are important to report. And of course, if outcomes are to be measured, then they have to be adjusted for the illness of patients and all methods used have to be completely transparent so that there is trust in those methods, or at least a good ability to understand how the results were generated.

TABLE 2: Translation of Measure Features into a Stakeholder Narrative

FOR CONSUMERS:	FOR PROVIDERS:
1. **Outcome Measure –** 20% rate of complications from delivery	1. **Level of Measurement –** This measures the Hospital's performance, not the individual Obstetricians
2. **Unit of Comparison –** Two star rating	2. **Measure has a Benchmark –** The national rate is 10%
3. **Measure has a Benchmark –** The national rate is 10%	3. **Outcome Measure –** 20% rate of complications from delivery
4. **Level of Measurement –** This measures the Hospital's performance, not the individual Obstetricians	4. **Risk Adjustment –** The measure is fully adjusted for the severity of the population treated
5. **Measure Focus is Condition, Procedure or Population –** The measure focuses on Deliveries	5. **Methodology is Transparent –** The complete definition of the measure and method for adjusting are published on www.hci3.org

Practically, reconciling these features isn't particularly difficult and Table 2 is an example of how a measure on rates of complications for deliveries could be reported to both consumers and physicians to satisfy their requirements.

Since outcome measures are important for consumers, it stands to reason that they should be readily available and widely reported, and yet they're not. There are, in fact, very few outcome measures reported publically, and mostly for hospitals. There are virtually no outcomes measures on physicians apart from the ones recently published by ProPublica and Consumers' Checkbook. However, our work for the past decade has uncovered that outcome measures can, in fact, be calculated for most conditions and procedures, at the facility or physician level, if what you look at are complications of care that could be avoided with optimal management of the patient.[16] This approach is also supported by significant research from many in the field.[17]

Prior Research and Findings

There's a reason why ProPublica and Consumers' Checkbook focused on complication rates of surgeons. In 2012, Professors Judith Hibbard of the University of Oregon, and Shoshanna Sofaer of the University of the City of New York, conducted several focus groups around the country to test ways in which pricing and quality information could be framed and presented to consumers in a way that would lead to making value-based decisions. They found that complication rates were the only measures that generated the desired effect because consumers intuitively understood that fewer complications could and should equate to lower costs of care.[18]

HCI[3] started developing measures of potentially avoidable complications (PACs) as part of its nationally recognized work on the PROMETHEUS Payment model, which was funded by charitable foundations (such as the Commonwealth Foundation and the Robert wood Johnson Foundation). Since then, these measures have been broadly used, researched, and analyzed.[19, 20, 21, 22] In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions,[23] they were adopted for various purposes, including the creation of related measures by other organizations.[24,25] Some commercial payers have used them as a means for tracking outcomes and for tiering providers for pay for performance programs.[26] In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement.[27] Identification of PACs has spurred provider innovation[28] for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability.[29] Some employers are also using measures of avoidable complications as public measures of quality[30] given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost.[18]

Accountability for and measurement of PACs occurs at the individual provider/practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 60% of its plan members with CAD incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or providers should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow-up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality, while also reducing costs.[31]

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality in the management of patients with a specific condition, illness or injury is likely to yield much improved outcomes for patients.[32] A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination. In addition, PAC measures can be used as a comprehensive outcome measure in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is no added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge the validity of its use.[33,34] Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased.[35] Importantly, in the current fee for service system, services for most PACs are rewarded by continued payment and hence, to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that occur and can cause harm to patients.[36]

The measurement of provider accountability for complications is not new. Medicare has instituted a number of penalty-based programs to curb some of them. For example, the "never events" payment policy eliminates any payment for a small number of egregious complications.[37] The more recent penalties for any readmission after a patient has been discharged following a hospitalization has shown some early promise.[38,39,40] And the implementation of new payment models such as bundled payments makes providers accountable for the full cost of any complication that might occur during the episode of care.[41,42]

There is therefore a solid body of evidence to support the use of complications as outcomes measures, and there are many measures of complications that are already being used.[43,44,45,46] Our work has continued to evolve to the point where we can meet the requirements of consumers and physicians in measuring rates of complications.[47]

Measuring & Reporting Comparative Rates Of Potentially Avoidable Complications

Broadly speaking, avoidable complications can be categorized into two types. Type 1 complications are directly related to the condition, illness or procedure being measured. For example, avoidable complications that are related to the management of heart failure may include hypotension, acute heart failure, or fluid and electrolyte disturbances. These complications can be avoided, in particular for patients under the age of 65, with active and continuous management and oversight of the patient. Similarly, avoidable complications that are related to a total knee replacement can include hemorrhage or a wound infection.

Type 2 complications are a broader set of patient safety-related complications such as drug-to-drug interactions, adverse effects to medication, line sepsis, deep vein thrombosis, phlebitis, falls, or other events that can occur when the system surrounding the patient fails.

Fieldwork has shown that physicians are far more likely to bristle at the inclusion of Type 2 complications because they feel they have less control over these events than Type 1 complications.[48] However, from the patient's perspective, all of these are complications, and almost all could be avoided if care were optimized.[49] From a more general policy perspective, if we are to move towards a truly patient-centered health care system, then the reduction of Type 2 complications are as important as Type 1 because they are more indicative of the lack of care coordination around the patient and of a functional system of care to support the patient through complete recovery.[50]

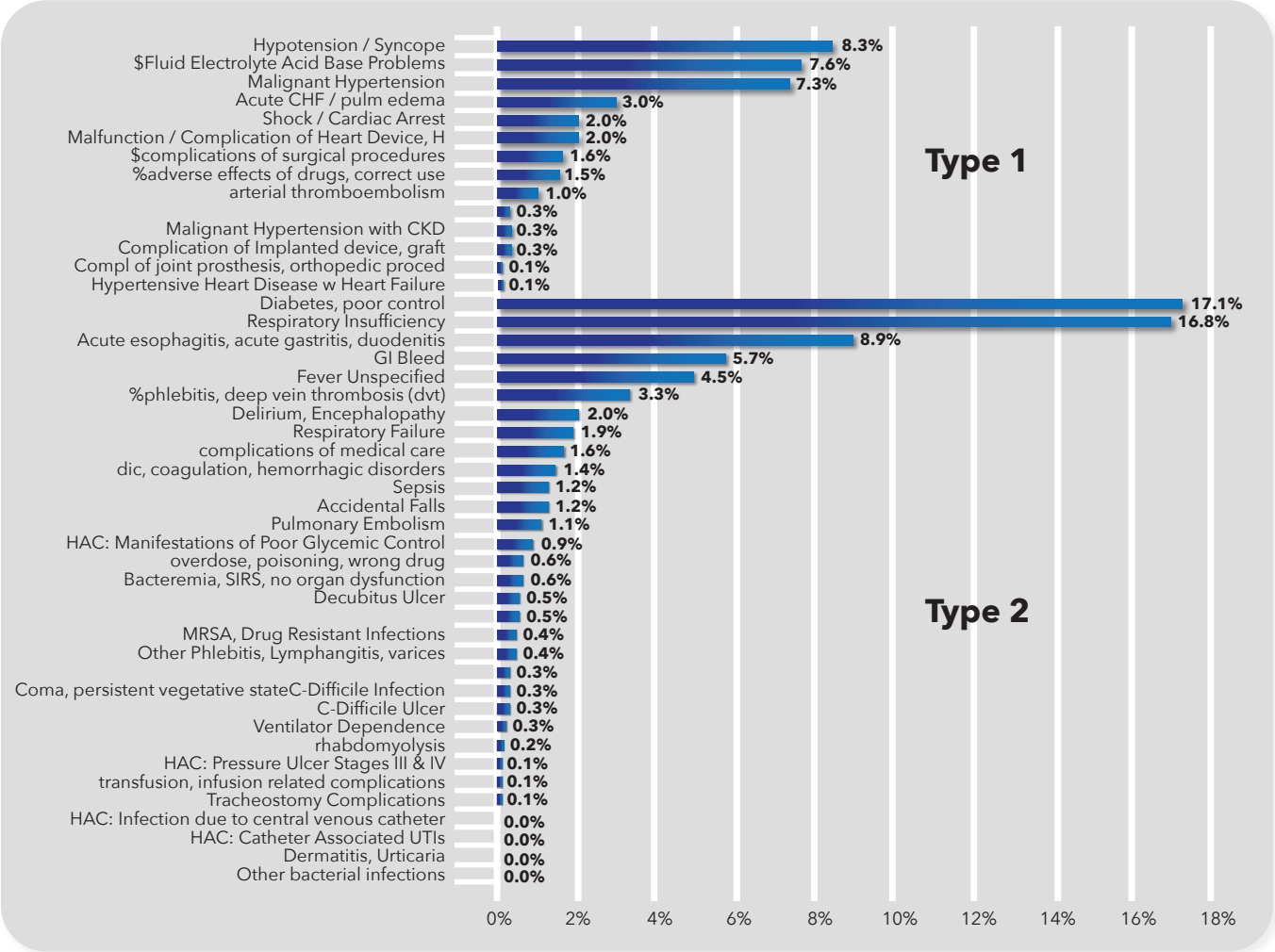
To illustrate the importance of measuring both types of complications, Table 3 summarizes the findings from a large commercial database on the frequency of potentially avoidable complications for patients with hypertension.

TABLE 3: Frequency of PACs in Hypertension Episodes

RELEVANT CASES	# UNIQUE PATIENTS	% OF TOTAL CASES
Count of Episodes	31,093	100.0%
Episodes w Any PAC	13,081	42.1%
Episodes with a PAC of Type 1	5,237	16.8%
Episodes with a PAC of Type 2	10,516	33.8%
Preventable Hospitalizations	1,100	3.5%

It's important to note that some patients who have hypertension can experience PACs of Type 1 as well as Type 2 during the course of a single episode-of-care. As such, these are not mutually exclusive and the frequency of these events clearly indicates the importance to measure them. And the following graph lists out the top PACs in each type that make up the majority of complications for patients with hypertension in the population studied.

FIGURE 1: Frequency of PACs for Hypertension Episodes



While the heterogeneity of Type 2 complications is evident from Figure 1, these represent the broader "system" failures that were decried in the series of reports on the quality of health care in America published by the Institute of Medicine at the turn of the century. The lack of comprehensive patient management, coordination between physicians around the care of the patient, and co-management of patient conditions creates the potential for adverse events. In this figure, for example, the single most frequent potentially avoidable complication is poor control of a patient's diabetes. This is also a CMS-defined HAC (Hospital Acquired Condition) that needs to be addressed by hospitals to avoid facing a penalty.[51] While this would also appear as a Type 1 PAC for diabetes episodes, the co-existence of diabetes and hypertension creates a need to ensure the proper management of both conditions by primary and specialty care physicians. By including these Type 2 PACs, the signal sent to physicians managing a patient's hypertension is that they should also work with the physicians managing the patient's diabetes to minimize the potential for negative events tied to the poor control of diabetes.

Of course, this figure also points out the importance to adjust for patient severity when comparing rates of avoidable complications. Clearly, patients who only have hypertension are very unlikely to get PACs for poor control of diabetes, but patients who have both hypertension and diabetes are likely to experience these PACs. Adjusting for the severity of patients should help account for the existence of multiple conditions in a single patient and the greater potential for PACs to occur. So let's examine the specific methods used to account for patient severity, create benchmarks for comparison, and ensure that the measures used reliably and fairly represent a physician's or facility's performance.

Description of Methods

Data The dataset used in all our analyses includes more than 3 million covered lives and over $25 billion in claims. Included are patient-level medical and pharmacy claims covering two years, from 2012 through 2014.

Episode Selection The episodes we selected for this analysis, which include a combination of elective procedures and chronic conditions, are listed in Table 4. We chose these episodes due to their high incidence in non-elderly privately insured populations and their impact on total costs. We also explicitly chose not to consider acute conditions and events since patients typically have little choice over their providers in these situations. As such, we believe these episodes are the most salient for these populations and those for which an individual would be likely to use provider-level quality information to make an informed treatment decision. That said, rates of potentially avoidable complications can be calculated for any episode of care when there is evidence of variability in performance and strength in the severity adjustment models. For example, while we had initially considered vaginal deliveries in our list of episodes, we excluded it from further analysis for two reasons. First, the risk adjustment models were very weak for these episodes and none of the facility-level reliability scores met acceptable levels. Second, based on ongoing analyses, we believe rates of cesarean sections are a more useful measure of a facility's overall quality of care for deliveries than rates of potentially avoidable complications of vaginal deliveries or C-sections

Triggering of Episodes and Inclusion and Exclusion Criteria
Episodes were identified or "triggered" based on the rules in the PROMETHEUS Analytics version 5.3[52] that take into consideration the combination of diagnostic and procedure codes contained in the patient claims. For procedures, episodes are triggered from an index hospitalization or outpatient claim, and condition episodes are triggered by a combination of ambulatory claims.

We excluded from the analysis episodes that failed to meet any of the following criteria:

1. Individuals less than 18 or more than 64 years of age;

2. Episodes in which the patient had a gap in enrollment of 30 days or more during the episode;

3. Episodes with total costs below the 1st percentile or above the 99th percentile.

4. Episodes that did not complete the predefined episode time period.

These purposeful exclusions prevented us from including incomplete episodes or those with claims that contained outlier codes or services. All condition episodes were annualized by taking the most recent 12 months of episode claims. The individual patient-episode is used as the base unit of analysis.

TABLE 4: Overview of Selected Episodes

EPISODE	EPISODES PER 1000*	% OF EPISODES WITH A PAC	AVG $ PER EPISODE	AVG PAC $ PER EPISODE
Chronic Conditions				
Asthma	28.8	40.2%	$769.03	$226.27
Coronary Artery Disease	10.4	47.0%	$1738.56	$413.50
Depression	26.0	24.3%	$1474.46	$469.15
Diabetes	27.8	59.6%	$1802.05	$622.50
Hypertension	88.3	31.6%	$973.81	$220.41
Low Back Pain	59.0	14.5%	$167.37	$12.05
Procedures				
Bariatric Surgery	1.9	45.7%	$19,598.42	$1,623.88
Coronary Angioplasty	1.9	48.6%	$21,913.00	$832.36
Knee Arthroscopy	5.4	13.1%	$8,034.63	$142.67
Lumbar Laminectomy	1.7	36.8%	$38,839.55	$1,450.81

*Episodes per 1000 plan members - Prevalence rate of episodes in the database

Occurrence of Potentially Avoidable Complications

Each episode definition includes codes for potentially avoidable complications (PACs). As claims get included in an episode, the costs of those claims get parsed, at the service line level when available, between typical and PAC costs. The overall rate of PACs in any episode can vary from none to over 90%. **The main outcome used in our analysis is a dichotomous measure (0=no, 1=yes) of the occurrence of at least one PAC during the episode period.**

Attribution of Episodes to Providers Episodes were attributed to providers or inpatient facilities according to attribution rules built into the PROMETHEUS Analytics. For procedures, the episode is attributed to the unique inpatient facility identifier listed on the index hospitalization claim. While procedures can also be attributed to a physician (for example the surgeon), that form of attribution relies extensively on the specialty code, when included, on the claims data, or an extensive mapping of the provider ID, such as a NPI, with the national NPI database. Because of the inconsistency with which these identifiers are included in claims, there is greater reliability in attributing procedures to facilities. For conditions, episodes are attributed to the provider with the highest count of office visits for the condition. Because providers or facilities with small volumes may provide unstable and/or unreliable estimates, we excluded from all provider-level analyses those that had fewer than 10 episodes in the data.

Defining the Measure Focus

As we discussed earlier, there are several features of quality measures that are essential to consumers and physicians. One is to clearly identify what the focus of the measure will be. Put simply, are we measuring the quality of multiple conditions at the same time, one condition at a time, or other combinations? To answer that question, let's first look at the results of the base analysis.

Descriptive Results As shown in Table 5, episodes vary widely in terms of the number of providers that take care of them and the number of episodes per provider. Moreover, there is significant variation in provider-specific PAC rates within each episode.

TABLE 5: Summary of Provider PAC Rates by Episode

EPISODE	# OF PROVIDERS	EPISODES PER PROVIDER		ACTUAL PAC RATES	
		AVERAGE	MIN - MAX	AVERAGE	MIN - MAX
Chronic Conditions					
Asthma	1,230	44	10 - 2,442	39%	0% - 100%
Coronary Artery Disease	457	42	10 - 1,157	45%	0% - 100%
Hypertension	3,657	57	10 - 3,128	34%	0% - 100%
Low Back Pain	3,280	42	10 - 1,448	13%	0% - 100%
Diabetes	1,659	33	10 - 889	60%	0% - 100%
Depression	1,151	28	10 - 481	23%	0% - 100%
Procedures					
Coronary Angioplasty	41	105	10 - 447	51%	32% - 80%
Bariatric Surgery	47	106	10 - 518	44%	15% - 85%
Knee Arthroscopy	374	39	10 - 523	14%	0% - 50%
Lumbar Laminectomy	58	52	10 - 270	36%	6% - 60%

Because of that variation, it might make sense to aggregate some of these conditions together, especially if they are proximate clinically (i.e. in the same clinical family). However, our analyses discourage combining PAC rates across different episodes into a single measure. That's because we looked at the relationships between providers' PAC rates for pairs of episodes and found little correlation. High correlations would suggest that providers' PAC rates are similar across episodes – those with low PAC rates tend to have low PAC rates for every episode and vice versa – and would support combining rates into a singular measure. Low correlations would indicate that PAC rates should be kept separate.

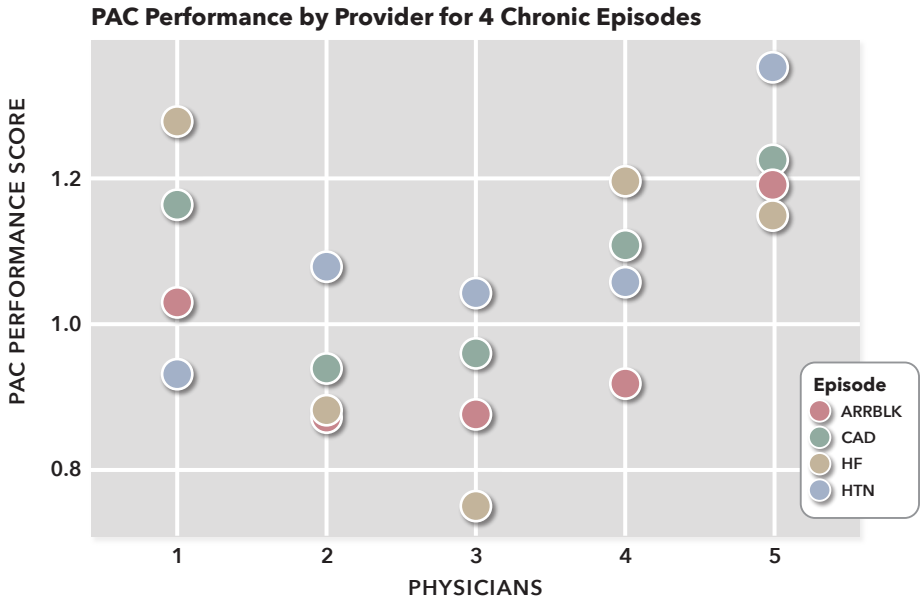
Table 6 shows the pairwise correlations coefficients between PAC rates for providers treating patients for each pair of chronic condition and procedure. With a few exceptions, the analysis shows that the associations between most combinations of conditions are generally weak or very weak. Moreover, these associations hold even when the PAC rates are risk adjusted. We therefore recommend that those who adopt this method for evaluating provider performance carefully test pairwise correlations before combining PAC rates across episodes. When the associations are weak, as they are here, PAC rates across episodes should be reported separately.

TABLE 6: Pairwise Correlations Between Provider PAC Rates

CHRONIC CONDITIONS	ASTHMA	CAD	HYPERTENSION	LOW BACK PAIN	DIABETES	DEPRESSION
Asthma	1.00					
CAD	0.32	1.00				
Hypertension	0.30	0.49	1.00			
Low Back Pain	0.13	0.26	0.17	1.00		
Diabetes	0.25	0.34	0.36	0.14	1.00	
Depression	0.31	0.22	0.31	0.24	0.33	1.00
PROCEDURES	**PCI**	**BARI SURG**	**KNEE ARTH**	**LUMBAR LAM**		
PCI	1.00					
Bari Surg	0.35	1.00				
Knee Arth	0.10	-0.04	1.00			
Lumbar Lam	0.01	-0.05	-0.06	1.00		

The importance of calculating the scores separately for each episode is underscored in Figure 2 in which we show the performance of physicians for related condition episodes. Some of the physicians seem to have a better performance than average, one (5) has a poorer performance across the board, but most have mixed results.

FIGURE 2: Comparison of PAC Rates by Physician and Episode



As a result of these correlation analyses, we recommend that the measure focus be the individual condition or procedure, and not combinations thereof. Let's now turn to another important feature, adjusting for patient severity.

Risk Adjustment Methods and Validation

In reporting any outcomes-based measure, it is important that the measure appropriately account for differences in the baseline health status of each provider's patient population. This ensures that providers are accurately and fairly compared in relation to their peers.

The risk adjustment models adjust for the following patient-related factors:

1. ***Patient demographics:*** Age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified for the patient.

2. ***Comorbidities:*** These are conditions or events that occurred prior to the start of the episode that could nonetheless have a potential impact on the patient's risk of having a PAC. These are universally applied across all episodes and identified from the diagnosis codes that appeared on an individual's claims prior to the start of the episode.

3. ***Episode Subtypes or Severity Markers:*** These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity), or severity of the illness itself (e.g., Hypertensive Heart Disease, Renovascular and other secondary hypertension), or the setting in which the procedure is performed (e.g., heart attack leading to an urgent PCI). Subtypes are specific to each unique episode.

All comorbidities and subtypes are identified prior to or at the very start of the episode to reduce the potential for gaming by upcoding claims.

Using these factors as covariates, we fit a logistic regression model to predict the probability of occurrence of a PAC during an episode. To prevent unstable coefficients, comorbidities and subtypes are included in the models as covariates only if they are present in at least 10 episodes. No further model building is performed once the initial models are built. The model preserves a very large group of covariates. This reflects a desire to explain as much variation as possible in the probability of having a PAC, without tailoring the predictors and introducing unnecessary bias This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of an episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode.

Of note, non-significant covariates in episode cost models cannot overly influence predicted outcomes, nor is much harm realized if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor. Separate models are fit for each episode and the predicted probabilities obtained from the models are used to construct the provider-level measures.

We validate our risk adjustment models using the split sample method. Specifically, episodes are randomly split into a development set (80% of episodes) and a validation set (20% of episodes).[53] The model is built on the development data set and then applied to the validation set. The outputs from these are then compared. We illustrate the strength of the models by reporting the Area Under the Curve (AUC) or c-statistics. The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. Values can range from 0.5 to 1.0, with 0.50 indicating that the model is no better than random prediction (i.e., the patient risk factors do not predict probability of occurrence of the outcome). Conversely, a c-statistic of 1.0 indicates perfect prediction (i.e., patients' outcomes can be predicted completely by their risk factors). Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.[54]

Comparisons of the AUC statistics are given in Table 7. Two important observations can be made about the AUC statistics from the table: 1) the models for all episodes have good discriminatory power and many are at or above the threshold at which models are considered strong, and 2) the statistics are virtually identical between the development and validation data sets. Overall, these results show that our models are sufficiently robust for risk adjusting PAC rates.

TABLE 7: Area Under the Curve (AUC) Comparisons by Episode

EPISODE	DEVELOPMENT SET	VALIDATION SET
Chronic Conditions		
Asthma	0.750	0.752
Coronary Artery Disease	0.801	0.799
Depression	0.800	0.801
Diabetes	0.839	0.835
Hypertension	0.814	0.811
Low Back Pain	0.790	0.778
Procedures		
Bariatric Surgery	0.724	0.684
Coronary Angioplasty	0.709	0.686
Knee Arthroscopy	0.716	0.700
Lumbar Laminectomy	0.734	0.690

We can therefore calculate risk adjusted PAC rates for providers – physicians and facilities – for specific episodes, and Table 8 includes an overview of the range of those rates for providers, by episode.

TABLE 8: Risk Adjusted PAC Rates Across Providers:

EPISODE	RSPR*	MIN - MAX
Chronic Conditions		
Asthma	38%	0 - 83%
Coronary Artery Disease	44%	0 - 89%
Depression	22%	0 - 86%
Diabetes	59%	0 - 99%
Hypertension	33%	0 - 98%
Low Back Pain	11%	0 - 66%
Procedures		
Bariatric Surgery	42%	16 - 68%
Coronary Angioplasty	52%	27 - 70%
Knee Arthroscopy	15%	0 - 59%
Lumbar Laminectomy	35%	9 - 67%

*Risk-Standardized PAC Rate

Calculating the results of a measure, however important that might be, is all for naught if the results of one provider aren't distinguishable from another. Earlier we saw that the second most important feature of quality measures for consumers is the ability to distinguish performance. In scientific language, that feature is referred to as reliability.

Reliability of PAC Rates As Outcomes Measures

Reliability is a measure that distinguishes between the signal (the extent of performance variation between entities that is due to true differences in performance) and statistical noise. It is important because it is an indicator of a measure's risk of misclassifying providers' performance. For example, high reliability would indicate a high performing provider or facility will most likely be classified as a high performer using the risk adjusted PAC rates; while low reliability would suggest they could be classified as low performing providers, when in fact they are high performers.

To test the reliability of risk adjusted PAC rates, we restricted the data to providers with at least 10 attributed episodes. We assessed the reliability of PAC rates using the beta-binomial method, which is applicable to measures of this type. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams and suggested by the National Quality Forum. This method yields an individual score for each provider or facility ranging from 0 to 1 with higher scores meaning better reliability.

There is no clear cut-off for an acceptable minimum level of reliability. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians.[55]

Details of the reliability analysis are shown in Table 9. For all the providers caring for chronic conditions, the median reliability scores were at or above the 0.70 threshold and, for the majority of providers – as evidenced by the inter-quartile ranges – they were above this number (middle columns of Table 9). For facilities that were attributed the procedural episodes, however, just one episode, bariatric surgery, achieved an average reliability score above 0.7. For the other procedures, most, it not all, facilities had scores under the threshold.

Because reliability scores provide a reasonable measure of assurance that PAC rates for certain providers are statistically distinguishable from those of the others, they can be used to determine minimum patient sample requirements for more accurately reporting an individual provider's performance for each episode. In the last two columns of Table 9 we show the minimum sample sizes for which all provider reliability scores exceed 0.7, and the percentage of all providers that met the criteria.

In reporting episode PAC rates for these groups of providers, we recommend only reporting the scores of those whose sample sizes exceed the minimum thresholds to achieve a reliability of 0.7 for each episode. Although, in some cases, many providers could be excluded from a comparative analysis, this approach does ensure that providers with small sample sizes are protected from being inaccurately mislabeled as a high or low performer.

TABLE 9: Summary of Reliability Scores by Episode

EPISODE	TOTAL # OF PROVIDERS	OVERALL RELIABILITY		POINT AT WHICH ALL SCORES > = 0.70	
		MEDIAN	IQR*	# EPISODES	% PROVIDERS OR FACILITIES
Chronic Conditions					
Asthma	1,231	0.79	0.69 - 0.89	20	50.1%
Hypertension	3,658	0.80	0.68 - 0.89	25	54.0%
CAD	458	0.73	0.62 - 0.83	25	36.5%
Low Back Pain	2,994	0.81	0.64 - 0.96	40	27.7%
Diabetes	1,660	0.73	0.63 - 0.83	25	34.0%
Depression	1,053	0.69	0.57 - 0.81	35	19.8%
Procedures					
PCI	40	0.47	0.28 - 0.62	185	12.5%
Bariatric Surg	47	0.87	0.80 - 0.93	25	80.9%
Knee Arth	374	0.05	0.03 - 0.21	^	0.0%
Lumbar Lam	58	0.50	0.32 - 0.72	80	27.6%

*Inter-quartile range (IQR)

In establishing minimum sample size requirements for PAC measures, it is important to point out that the reliability calculations will be determined by the unique data set on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset or population would apply to another.

Provider Comparisons and Benchmarks

In this section, we show how to calculate provider PAC rates for comparison purposes and how to translate these into information that is understandable for consumers and important to physicians. To construct measures that allow for direct and meaningful comparisons between individual providers, risk-standardized PAC rates (RSPR) are used. This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures such as for mortality and for readmission rates. The calculation of the RSPR is as follows:

- For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.
- Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PAC rates for the provider.
- The observed sum is then divided by the summed probabilities (O/E). This number yields a performance ratio indicating whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1).
- This ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted PAC rate would be if its patient population was reflective of the of the overall community. The formula for this calculation is as follows:

$$\text{Risk Standardized PAC Rate}_j = \left(\frac{\sum \text{Episodes with a PACs}_{i,j}}{\sum \text{Probability of a PAC}_{i,j}}\right) * \left(\frac{\sum \text{Probability of a PAC}_{i,j}}{\text{Total \# of episodes}}\right)$$

Where individual i is attributed to unit of analysis j (e.g. physician, facility, etc.)

The application of the risk standardized PAC rates for reporting purposes should be tailored to the audience that will use the information. While a risk standardized PAC rate may be useful for the providers themselves, they may be hard to interpret for most consumers. Instead, consumers need a way to simply, yet intelligently, identify a high quality provider over a low quality provider. A more useful way to do this, for instance, is to categorize providers into groups based on their PAC rates in relation to some benchmark, such as the average.

In order to facilitate the use of PAC rates by consumers, we show, using asthma and hypertension episodes as examples, a simple way of categorizing providers into different levels of PAC rates. To start, we only included providers that met the minimum sample size requirements that were established in the reliability analysis above. In order for higher scores to translate into lower PAC rates, we also subtracted the risk-standardized rates from 1.

Providers for each episode were split into three categories: below average, average, and above average. Inclusion in the above and below average categories was based on whether a provider was above or below one standard deviation of the average risk-standardized PAC rate for all providers. Providers with PAC rates more than one standard deviation above the average were labeled as "below average." Similarly, providers with PAC rates more than one standard deviation below the average were labeled as "above average." Providers within one standard deviation of the average were considered "average" performers.

The breakdown of providers across performance categories for two sample episodes is shown in Table 10. Between two-thirds and three-quarters of providers are labeled as having average PAC rates. Because higher PAC rates equate to lower performance, the average risk standardized PAC rates decrease with higher performance categories.

TABLE 10: Provider PAC Performance

	BELOW AVERAGE (HIGH PAC RATE)	**AVERAGE** (AVERAGE PAC RATE)	**ABOVE AVERAGE** (LOW PAC RATE)
Asthma			
% of Providers	17%	67%	16%
Average RSPR	56%	38%	19%
Range (Min - Max)	51 - 73%	26 - 51%	0 - 25%
Hypertension			
% of Providers	14%	72%	14%
Average RSPR	49%	33%	19%
Range (Min - Max)	43 - 93%	23 - 43%	5 - 23%

A diagram showing the complete distribution of physicians' PAC performance is provided in Figures 3 and 4, with physicians above the blue zone having worse performance and those below having better performance.

FIGURE 3: Distribution of RSPR by Physician for Asthma

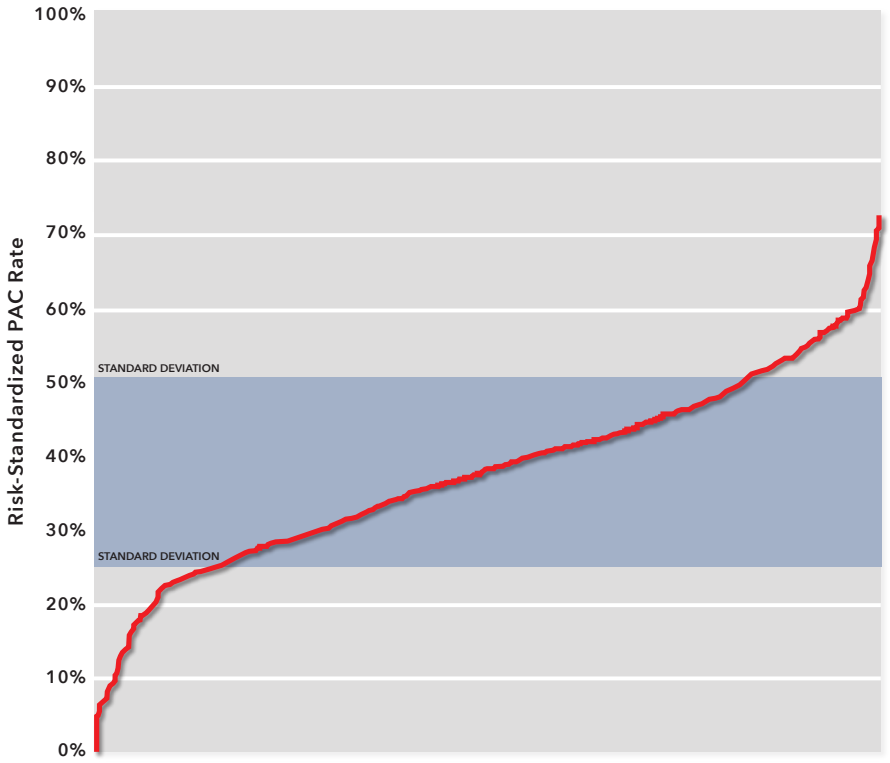
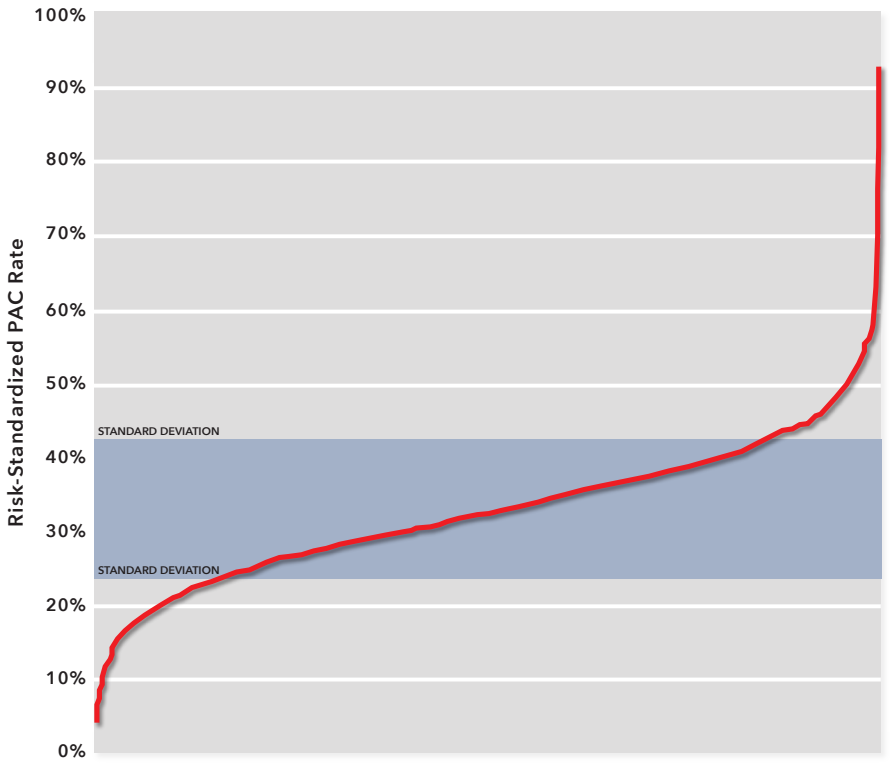


FIGURE 4: Distribution of RSPR by Physician for Hypertension



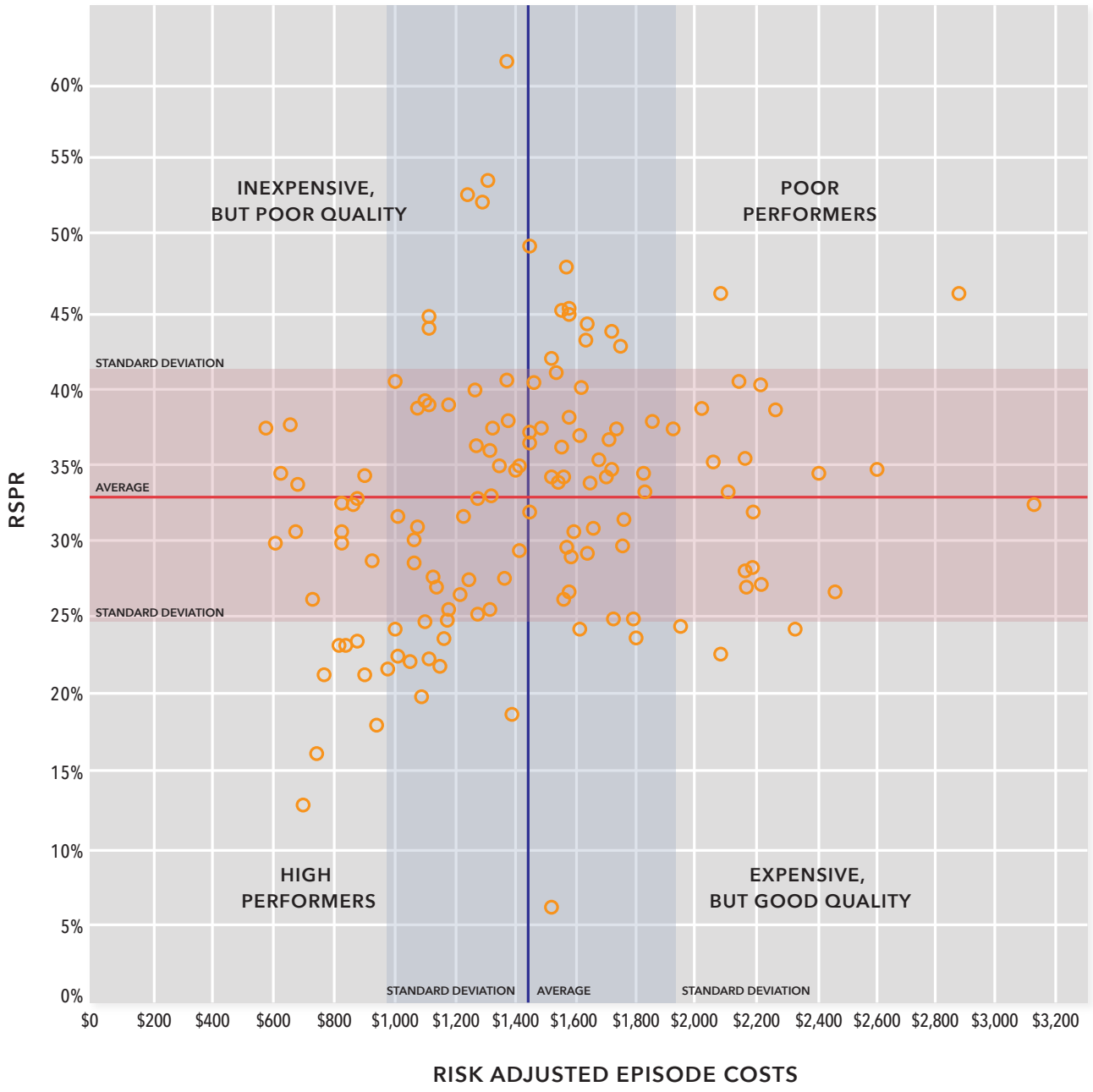
The Path Ahead

With the introduction of a risk-standardized rate of potentially avoidable complications that can be measured at the individual condition or procedure level, and applied to individual physicians or facilities, we have the potential to enter into a new era of quality of care transparency. The data source for these comprehensive outcome measures is claims, which are plentiful and far easier to access than medical record data. In fact, the growing availability of Medicare claims data by CMS, and commercial insurance and Medicaid data by all-payer claims databases (APCD), provide a unique opportunity for a potential "big bang." There are, however, some barriers to accessing those data.

In some instances, the APCD data stewards are governed by committees that are dominated by physicians, and full and complete transparency of quality of care is still a scary proposition for many. As a result, the ability to freely use APCDs to publish measures such as risk-standardized PAC rates is likely to be a state by state fight, and one that we are ready to wage.

Conversely, private sector payers and state Medicaid agencies could leverage our work to make their own push for more complete quality transparency. Our Open Source episode definitions already provide all of the information on how to construct and define episodes of care and identify potentially avoidable complications. Adjusting for severity and testing for reliability are relatively straightforward statistical methods that we will make widely available to facilitate these calculations. Further, as we show in Figure 5, by combining RSPR with Severity-adjusted Costs of Episodes, there is a unique opportunity to help consumers define value. For example, in Figure 5, physicians in the lower left quadrant have both low total episode costs and low rates of complications, while those in the upper right quadrant have high PAC rates and high episode costs.

FIGURE 5: Distribution of Physicians Based on RSPR and Episode Costs



[1]How to Shop for Medical Care 101 – Jun 2015: No more "Buying Blind" & Tackling Rising Healthcare Costs with Healthcare Transparency http://thinkhealth.priorityhealth.com/tackling-rising-health-care-costs-with-health-care-transparency/ - Accessed Nov 1 2015.

[2]Suzanne DelBanco and Francois deBrantes – July 2015. Report Card on State Price Transparency Laws: http://www.catalyzepaymentreform.org/images/documents/2015_Report_PriceTransLaws_06.pdf - Accessed Oct 30 2015.

[3]Dan Mangan CNBC 3 Feb 2015. Out of Pocket Healthcare Costs rise, but prices often remain hidden: http://www.cnbc.com/2015/02/02/as-out-of-pocket-health-costs-rise-prices-often-remain-hidden.html - Accessed Nov 1 2015

[4]Gregory and Stuart, Paul and Robert (2004). Comparing Economic Systems in the Twenty-First Century, Seventh Edition. George Hoffman. p. 538. ISBN 0-618-26181-8. "Market Economy: Economy in which fundamentals of supply also demand provide signals regarding resource utilization."

[5]Burns LR. "India's Healthcare Industry" Cambridge University Press, 2014

[6]Ari Mwachofi, Assaf F. Al-Assaf. Health Care Market Deviations from the Ideal Market. Sultan Qaboos Univ Med J. 2011 August; 11(3): 328–337. Published online 2011 August 15. PMCID: PMC3210041

[7]Hibbard J, Greene J. "What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs." Health Affairs. University of Chicago. March 1, 2013.

[8]Bennett L, de Brantes F, DiLorenzo J, McGuire W. HCI3 Dec 2014: "State Report Card on Transparency of Physician Quality Information. http://www.hci3.org/content/physician-quality-transparency-report-2014 - Accessed Nov 1 2015.

[9]The Quality Measures that Matter: Action Brief: http://www.catalyzepaymentreform.org/images/Action_Brief_QualityMeasure_Final_10.29.15.pdf - Accessed Oct 30 2015.

[10]CPR Employer-Purchase Guide to Quality Measure Selection: http://www.catalyzepaymentreform.org/images/CPR_Employer-Purchaser_Guide_to_Quality_Measure_Selection_2015-10-23.pdf - Accessed Oct 30 2015.

[11]Pro-Publica Surgeon Scorecard – July 2015: https://projects.propublica.org/surgeons/ - Accessed Nob 1 2015

[12]Consumer Checkbook - Surgeon's Ratings: http://www.checkbook.org/surgeonratings/?action=article&articleID=how-big-are-the-differences-among-surgeons - Accessed Nov 1 2015

[13]Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

[14]Jha, A. "The ProPublica Report Card: A Step in the Right Direction." An Ounce of Evidence: Health Policy. Web.

[15]Friedberg MW, Pronovost PJ, Shahian DM, et.al. A Methodological Critique of the ProPublica Surgeon Scorecard Aug 2015: http://www.rand.org/pubs/perspectives/PE170.html - Accessed Nov 1 2015.

[16]Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

[17]James JT. A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care. J Patient Saf, 2013 (Sept); 9 (3): 122-128.

[18]Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

[19]de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

[20]de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

[21]François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability –The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

[22]deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

[23]Diana Manos Healthcare IT News. NQF endorses four new patient quality care measures http://www.healthcareitnews.com/news/nqf-endorses-four-new-patient-quality-care-measures - Accessed Nov 1 2015.

[24]National Quality Forum – Quality Positioning System: http://bit.ly/1MAR7Im - Accessed Nov 1 2015.

See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

[25]Robert Wood Johnson Foundation: http://www.rwjf.org/en/library/research/2011/02/prometheus-payment/the-fundamentals-how-the-model-works.html - Accessed Nov 1 2015.

[26]Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

[27]Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov, June 2012. Web. CAPLERS http://bit.ly/1FGYk0M

[28]2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

[29]Micaela P. McVary. The Prometheus Model: Bringing Healthcare into the Next Decade. Annals of Health Law Advance Directive, 2010 (Spring); 19: 274-284.

[30]Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/resources/reducing-costs/healthcare-incentives-payment-pilot-hipp/ - Accessed Nov 1 2015.

[31]de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, - Accessed October 30 2015.

[32]Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

[33]Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web

[34]Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

[35]Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

[36]Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

[37]Arnold Milstein. Ending Extra Payment for "Never Events" – Stronger Incentives for Patients' Safety NEJM 2009: 360:2388-2390 DOI: 10.1056/NEJMp0809125. http://www.nejm.org/doi/full/10.1056/NEJMp0809125

[38]US News – Oct 19, 2015: Medicare Incentives and Penalties: Critical Lessons Learned: http://health.usnews.com/health-news/hospital-of-tomorrow/articles/2015/10/19/medicare-incentives-and-penalties-critical-lessons-learned - Accessed Nov 1 2015

[39]Best way to Run a Hospital Readmission Program. https://www.healthcatalyst.com/healthcare-data-warehouse-hospital-readmissions-reduction - Accessed Nov 1 2015

[40]AHRQ: 2013 Annual Hospital-Acquired Condition Rate and Estimates of Cost Savings and Deaths Averted from 2010 to 2013. http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/hacrate2013.pdf - Accessed Nov 3 2015.

[41]Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

[42]Hussey PS, Sorbero ME, Mehrotra A et al. Episode-based Performance Measurement and Payment: Making it a Reality. Health Aff (Millwood). 2009; 28(5):1406-17; 10.1377/hlthaff.28.5.1406.

[43]S.F. Jencks, M.V. Williams, and E.A. Coleman: Rehospitalizations among Patients in the Medicare Fee-for-Service Program. New England Journal of Medicine 2009; 360: 1418-1428.

[44]Michael R. Eber; Ramanan Laxminarayan; Eli N. Perencevich; Anup Malani. Clinical and Economic Outcomes Attributable to Health Care-Associated Sepsis and Pneumonia. Arch Intern Med, 2010; 170 (4): 347-353

[45]J. Hughes, "Identifying Potentially Preventable Complications Using a Present on Admission Indicator," Health Care Financing Review 27, no. 3 (2006): 63-82.

[46]David Blumenthal: Reflecting on Health Reform: Good News! High Performance in Action. The Commonwealth Fund Blog, April 16, 2014. http://www.commonwealthfund.org/publications/blog/2014/apr/drop-in-infections - Accessed Nov 1 2015.

[47]HCI3: Measuring Provider Performance: http://www.hci3.org/content/measuring-provider-performance - Accessed Nov 1 2015

[48]NQF Draft Report: https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=80820 - Page 24, Oct 2015.

[49]The Commonwealth Fund. March 2011. OSF HealthCare: Promoting Patient Safety Through Education and Staff Engagement. http://www.commonwealthfund.org/publications/case-studies/2011/mar/osf-healthcare

[50]Wachter RM, Pronovost PJ, and Shekelle PG. Editorial. Strategies to Improve Patient Safety: The Evidence Base Matures. Annals of Internal Medicine, 2013; 158(5): 350-352.

[51]CMS defined Hospital-Acquired Conditions: https://www.cms.gov/medicare/medicare-fee-for-service-payment/hospitalacqcond/hospital-acquired_conditions.html - Accessed Nov 1 2015.

[52]For more information on the episode definitions and trigger rules, please refer to the HCI3 web site (www.hci3.org).

[53]Iezzoni LI, ed. Risk Adjustment for Measuring Health Care Outcomes, 3rd ed. Chicago, IL: Health Administration Press, 2003.

[54]Hosmer DW, Lemeshow S. Applied Logistic Regression (2nd Edition). New York, NY: John Wiley & Sons; 2000.

[55]Adams JL. The Reliability of Provider Profiling: A Tutorial. 2010. Rand Corporation: http://www.rand.org/pubs/technical_reports/TR653.html - Accessed Nov 4 2015

We hope that with the publication of this White Paper, and given the continued poor performance of most states on the transparency of the quality of physician care, States will take a more pro-active role to provide consumers with the information they need to make informed choices.

13 SUGAR STREET
NEWTOWN, CT 06470
EMAIL: INFO@HCI3.ORG
WWW.HCI3.ORG