

PACs and Severity Adjustment Fact Sheet

Background:

Potentially avoidable complications (PAC) measures are comprehensive measures of adverse events that occur either due to errors directly related to the index condition and under the control of the servicing physician, such as hemorrhage following bypass procedure (type 1 PACs), or due to patient safety failures, such as pressure ulcers and central line infections (type 2 PACs). While some of these adverse events cannot be avoided in the best of hands and within the best of systems, PAC measures are an attempt to measure all harms that occur to the patient within the healthcare system.

Measures associated to potentially avoidable complications (PACs) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures^{i,ii,iii}. In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures^{iv}. Some commercial payers have used them as a means for tracking outcomes^v and for tiering providers for pay for performance programs^{vi}. 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^{vii}. Identification of PACs has spurred provider innovation^{viii} for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability^{ix}. Some employers are also using measures of avoidable complications as public measures of quality^x given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost^{xi}.

A recent Health Affairs article highlights the inconsistency of national rating systems, even for hospitals with significant case volume, suggesting that overall composite scores, or even scores on specific procedures or acute events are highly dependent on the methodology used to assess and score quality outcomes^{xii}. In addition, several recent papers highlight that many existing quality measures collected and reported on Medicare's Hospital Compare website, as well as through national quality improvement programs, fail to differentiate providers based on quality scores^{xiii}. As such, attempting to show any correlation between currently reported measures, whether individual measures or composites, and these proposed measures of potentially avoidable complications, is somewhat futile.

Instead, in this application, our approach is to highlight the face validity of the measures and demonstrate they can sufficiently differentiate performance between providers and hospitals, after appropriate risk-adjustment. But first, it might be helpful to trace back the development of PACs.

Development of PACs since 2007:

In 2006 and 2007, under a grant from The Commonwealth Fund, we launched a project to help parse out, within an episode of medical care, the variability that was explainable (e.g. due to the severity of the patient), from the variability that wasn't explainable, and in particular was completely unwarranted. Prior research, in particular from Dartmouth, had shown that there was a lot of variability due to patient preference, supply of services and provider preference. Yet the IOM reports on the Quality of Care in America also showed alarming rates of “defects” – misuse of services – that could cause patient harm.

With the help of clinical experts assembled by clinical domain, we started delineating the boundaries of episodes and to express them in terms of both clinical guidelines, but also diagnosis and procedure codes in order to analyze claims data and determine the extent to which there were measurable negative (and costly) events occurring, and their frequency. The proceeds of that initial work was published in a report ^{xiv} and became the basis for the development of the PROMETHEUS Payment model. What we found was, and remains, astounding. Not only was the frequency of avoidable complications high, the variability of avoidable complications by provider and across the country was significant^{xv}.

Since then, variations on this work have been espoused by others. For example, 3M came out with its Potentially Preventable Admissions and Readmissions^{xvi}, which constitutes one of the more costly subsets of HCI³'s PACs. CMS, through the Medicare program, introduced penalties for all-cause 30-day readmissions^{xvii}, which is the broadest attempt at capturing PACs and widening the scope of what we consider to be appropriately defined avoidable complications linked to a specific episode of medical care.

In early 2011, the NQF endorsed several PAC measures as comprehensive outcome measures, thus validating the many years of research and empirical analyses about the importance of PACs. Today, the development and refinement work continues, with observations from very large datasets, different types of payers, and many more episodes of care than originally defined.

Face Validity:

The development of PAC definitions has continuously involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focus on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help define and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (<http://www.hci3.org/content/clinical-working-group-contributors>). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures

(<http://www.hci3.org/content/using-ecrs-providers>).

Beyond the up front work performed by clinical experts, the validity of the measures have been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues^{xviii} tried to assess the impact of presenting information about price and quality of certain providers in influencing the decision of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer “defects” – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)^{xxix}, Hospital Inpatient Quality Reporting measures^{xxx}, Avoidable Readmissions^{xxixxxii}, AHRQ defined patient safety indicators (PSIs)^{xxiii}, NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates^{xxiv}.

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.), and the parallels between this measure and other measures that are in widespread use, we believe this demonstrates that the measure has strong face validity.

Validity of Risk Adjustment Techniques:

Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, such as patient demographics, comorbidities, severity of illness, complexity of the procedure etc., the remaining variance in PAC rates are probably due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window. Because our measure of PACs (potentially avoidable complications) is based on the occurrence of one or more PACs during the episode, the severity-adjustment models are intended to predict the probability that at least one PAC occurred during an episode given the patient's combination of risk factors and comorbidities. To avoid creating perverse incentives, the models consider only those comorbidities and indicators of episode severity that are present at the start of the episode.

Unit of Analysis:

The unit of analysis is the individual episode. Episodes above and below the 1st and 99th percentiles based on episode costs are excluded from the analysis to avoid including any potential false positive or unusual episodes that could impact the results. Also excluded are episodes in which the patient is less than 18, individuals within an enrollment gap at any time during the episode, and incomplete episodes, whose entire episode time window is not captured within the data.

Dependent Variable:

The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:

A number of patient-related “risk factors” or covariates are included in the models:

- *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.
- *Comorbidities:* These are conditions or events that occurred prior to the start of the episode that can 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 a member’s historical. The list of all comorbidities that can be potentially identified are listed in the workbook attached to each measure in the submission.
- *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). Subtypes are specific to each unique episode.

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods:

We use logistic regression to model the probability of at least one PAC occurring during the 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 conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the models be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. 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. Non-significant covariates in episode cost models can not 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.

Some limitations include:

1. Models are only as good as the data.
2. We're using claims data from specific populations to estimate PAC rates. These estimates may not apply to all populations.
3. Claims data are messy with incomplete or incorrect diagnosis codes being used and there is limited clinical information that can only be obtained through patient chart reviews such as information on body weight, BP, smoking status etc.
4. Socio-economic and other factors such as educational level, family support, access to healthcare etc., which may impact costs are not available in most administrative databases but are very important for severity adjustment.
5. Laboratory data is not available to us – it has been shown to add a lot of value to administrative data for risk-adjustment purposes to assess the severity of a patient's disease.

Model Validation

To determine if our risk models are valid, we used a split sample method^{xxv} to divide the patient sample into: 1) the Model building dataset (80% of sample) and 2) the Test dataset (20% of sample). The model was built using logistic regression techniques on the first dataset and then the coefficients from the development model were tested in the second dataset. Area of the curve (AUC) statistics were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy. Results of these tests are reported for each measure in the submission.

Variation in Provider Performance:

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes between signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report “The Reliability of Provider Profiling: A Tutorial” by J.L. Adams^{xxvi}.

To directly compare PAC rates across providers or facilities while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcome measures.^{iv} For each provider or facility, the ratio of observed

attributed episodes with PACs to the expected number of attributed episodes with PACs, given the patient's risk factors, and estimated from the risk-adjustment model^{xxvii} was calculated. This number yielded whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers or facilities to obtain the risk-standardized PAC rate for the provider or facility. Because providers or facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance.

Model Validation

Our analysis compared risk-standardized PAC rates across providers or facilities. We show several descriptive statistics including the range in PAC rates, deviations from the average rate, etc. in each measure submission.

Reliability Testing:

Reliability testing was done on the risk-standardized PAC rates (RSPR). Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

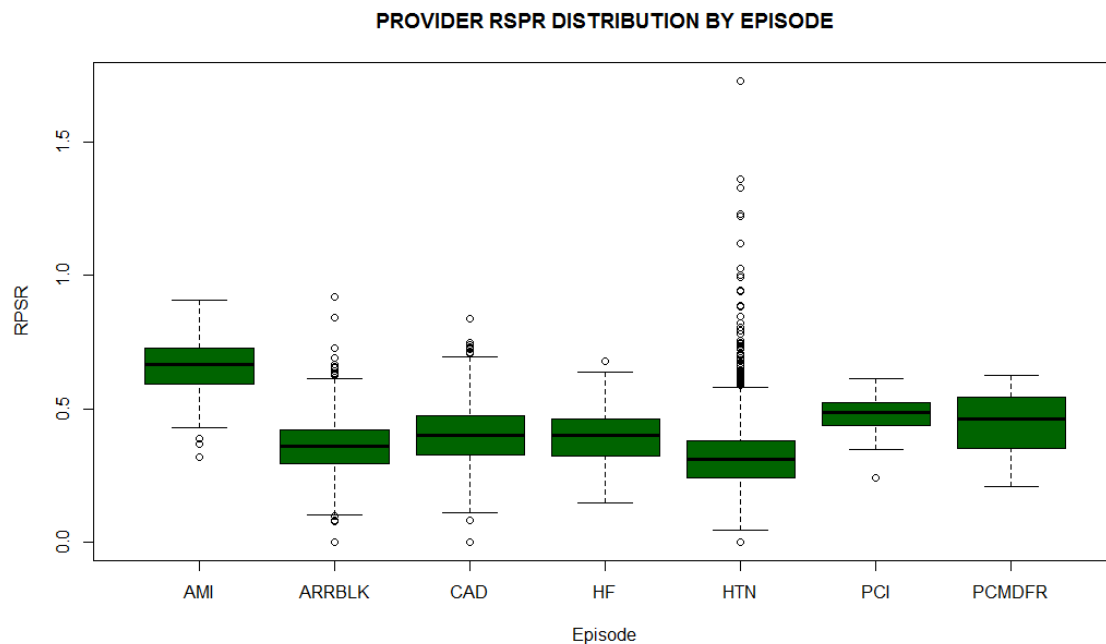
There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

The table below shows the minimum sample size required for the various PAC measures to achieve a Reliability score of >0.7 in the dataset studied. On an average, higher sample sizes are required for procedural episodes because of the lower variability in rates of complications, and vice versa for chronic condition episodes (lower sample size requirements because of much higher variability in rates of complications)

Episodes	Sample Size for Absolute Reliability >0.7	Sample Size for Median Reliability >0.7
PCI	185	175
CAD	21	10
Heart Failure	31	25
Hypertension	23	10
Arrhythmia / Heart Block	30	25

Variation in Risk-Adjusted Provider Performance Scores by Episode

The chart below shows the distribution of provider performance scores for each episode in the dataset analyzed. Note, RSPR refers to the risk-standardized PAC rate.



Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset 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 would apply to another.

References:

- ⁱ deBrantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. *Health Serv Res.* 2010 Dec; 45(6 Pt2): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x
- ⁱⁱ Joynt KE, Gawande AA, Orav EJ, and Jha AK. Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients. *JAMA* 2013; 309(24): 2572-2578. doi: 10.1001/jama.2013.7103.
- ⁱⁱⁱ James JT. A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care. *J Patient Saf*, 2013 (Sept); 9 (3): 122-128.

^{iv} 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>

^v The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop 2010 ... By Roundtable on Evidence-Based Medicine, LeighAnne Olsen, Pierre L. Yong, Roundtable on Value & Science-Driven Health Care, Institute of Medicine. Online version: <http://bit.ly/1FN9PTTr>

^{vi} Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

^{vii} CAPLERS <http://bit.ly/1FGYk0M>

^{viii} 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5.
<http://www.bundledpaymentsummit.com/agenda/day1.html>

^{ix} Micaela P. McVary. The Prometheus Model: Bringing Healthcare into the Next Decade. *Annals of Health Law Advance Directive*, 2010 (Spring); 19: 274-284.

^x Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP):
<http://www.cbghhealth.org/resources/reducing-costs/healthcare-incentives-payment-pilot-hipp/>

^{xi} 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)*. 2012 Mar; 31(3):560-8. doi: 10.1377/hlthaff.2011.1168.

^{xii} Austin JM, Jha AK, Romano PS et.al. National Hospital Ratings Systems share few common scores and may generate confusion instead of clarity. *Health Affairs*, 2015; 34 (3): 423-430. doi: 10.1377/hlthaff.2014.0201. online version:
<http://content.healthaffairs.org/content/34/3/423.full.html>

^{xiii} Kyan C. Safavi M.D., M.B.A.*, Feng Dai Ph.D., Todd A. Gilbertsen B.S. and Robert B. Schonberger M.D., M.A. Variation in Surgical Quality Measure Adherence within Hospital Referral Regions: Do Publicly Reported Surgical Quality Measures Distinguish among Hospitals That Patients Are Likely to Compare? *Health Services Research*, 2014 (August); 49 (4): 1108–1120.

^{xiv} See <http://www.hci3.org/content/evidence-informed-case-rates-new-health-care-payment-model>

^{xv} See <http://www.hci3.org/content/episode-care-analysis-reveals-sources-variations-costs>

^{xvi} 3M-Potentially Preventable Readmissions:
<http://multimedia.3m.com/mws/media/7842130/3m-pop-health-ppe-ebook.pdf?&fn=Preventables%20eBook.pdf>

^{xvii} MedPac 2013. Refining the Hospital Readmissions Reduction Program. Rep. no. 4. Vol. 4. Washington DC: <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html>

^{xviii} Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

^{xxix} CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

^{xx} CMS operated Hospital Inpatient Quality Reporting Program: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html>

^{xxi} Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. *N Engl J Med* 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

^{xxii} Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. *Health Affairs*, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

^{xxiii} Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

^{xxiv} NQF endorsed measures: Quality Positioning System: <http://bit.ly/1E5ZdP7>

^{xxv} Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press, 2003.

^{xxvi} Adams JL. The Reliability of Provider Profiling: A Tutorial. 2010. Rand Corporation.

^{xxvii} Rathouz PJ and Rastogi A. *Linear Regression in Medical Research*. In Bailar JC and Hoaglin DC Eds. *Medical Uses of Statistics*, 3rd edition, Chapter 10, NEJM Press 2009, New Jersey: John Wiley & Sons Inc. 2009. ISBN: 978-0-470-43952-4 (cloth) and ISBN: 978-0-470-43953-1 (pbk.)