NATIONAL QUALITY FORUM

Memo

TO: Cardiovascular Standing Committee

FR: NQF Staff

RE: Post-Comment Call to Reconsider HCI3 Measures

DA: January 28, 2016

Purpose of the Call

The Cardiovascular Standing Committee will meet via conference call on Thursday, January 28, 2016 from 3:00pm-5:00pm (ET). The purpose of this call is to:

- Review and discuss comments pertaining to the HCI3 measures received during the post-evaluation public and member comment period;
- Provide input on proposed responses to the post-evaluation comments pertaining to the HCI3 measures;
- Discuss and re-vote on one measure that did not reach consensus; and
- Reconsider five measures using the NQF measure endorsement criteria algorithm.

Standing Committee Actions

1. Review this briefing memo and Draft Report.
2. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see Comment Table and additional documents included with the call materials).
3. Be prepared to provide feedback and input on proposed post-evaluation comment responses.

Conference Call Information

Please use the following information to access the conference call line and webinar:

Speaker dial-in #: 1-(877) 298-1950
Web Link: http://nqf.commpartners.com/se/Rd/Mt.aspx?141191
Registration Link: http://nqf.commpartners.com/se/Rd/Rg.aspx?141191

Background

During the 2-day in-person meeting, the 20-member Cardiovascular Standing Committee evaluated six measures submitted by the Health Care Incentives Improvement Institute (HCI3). The purpose of these measures is to identify the magnitude of potential avoidable complications (PACs) and the cause of the most frequent and costly complications in order to focus on reducing those PACs and ultimately improve patient outcomes.

PACs are defined as Type 1 PACs and Type 2 PACs. Type 1 PACs are complications directly related to the index condition. Patients are considered to have a Type 1 PAC if they receive services during the episode time window for complications directly related to coronary artery disease (CAD), heart failure (HF), hypertension (HTN), arrhythmias, pacemaker/defibrillator implantation, or angioplasty. Examples of Type 1 complications include hypotension, cardiac arrest, fluid and electrolyte disturbances, wound infection, etc. Type 2 PACS are considered

patient safety failures such as sepsis, infections, phlebitis, DVT, pressure ulcers, etc. Patients are considered to have a Type 2 PAC if they receive services during the episode time window for any complications related to patient safety failures.

During the in-person meeting the Committee agreed the Type 1 PACs were more directly related to the measure conditions but some Committee members expressed substantial concern that the Type 2 PACs were too broad and that I individual clinicians would be held responsible for PACs unrelated to the management of the index condition. Committee members also expressed concern that there was no evidence or rationale provided to support the one-year time frame for the four condition specific measures (#2740, #2749, #2747, and #2748) and the selection of the Type 2 PACs. The Committee's greatest concern was that the four measures are specified at the clinician level, rather than the facility level, which many asserted was more appropriate. The two procedure specific measures are specified at the facility level (#2751 and #2752). The Committee expressed greater comfort in attributing any complications associated with procedures at the facility level of analysis, but remained concerned that the facility would be held responsible for Type 2 PACs unrelated to these procedures.

The Committee did not reach consensus on one measure (#2751) and five measures were not recommended for endorsement (#2740, #2749, #2747, #2748, and #2752). NQF allows for requests for reconsideration by measure developers. As detailed below, HCI3 has prepared a detailed memo to accompany their reconsideration requests for the five measures that were not recommended by the Standing Committee.

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments prior to the evaluation of the measures via an online tool located on the project webpage. Third, NQF opens a 30-day comment period to both members and the public after measures have been evaluated by the full committee and once a report of the proceedings has been drafted.

Pre-evaluation comments

The pre-evaluation comment period was open from July 29, 2015 to August 12, 2015 for the six HCI3 measures under review. One pre-evaluation comment was received in support of one of the six measures and provided to the Committee prior to their initial deliberations held during the in-person meeting.

Post-evaluation comments

The Draft Report went out for Public and Member comment October 23, 2015 to November 23, 2015. During this commenting period, NQF received 13 comments from 2 member organizations and 1 public representative:

Consumers – 0 Professional – 1

Purchasers – 0	Health Plans – 1
Providers – 0	QMRI – 0
Supplier and Industry – 0	Public & Community Health - 1

In order to facilitate discussion, the majority of the post-evaluation comments have been categorized into major topic areas or themes. NQF staff has proposed draft responses for the Committee to consider. Although all comments and proposed responses are subject to discussion, we will not necessarily discuss each comment and response on the post-comment call. The Committee will be asked if they would like to further discuss any comments received before determining the final Committee responses.

NQF has included all of the comments that were received (both pre- and post-evaluation) in the Comment Table. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and—for the post-evaluation comments—draft responses for the Committee's consideration. Please refer to this comment table to view and consider the individual comments received and the proposed responses to each.

Comments and their Disposition

One theme was identified in the post-evaluation comments: commenters agreed with the Committee's recommendations.

Theme 1 – Agree with Committee's recommendations

Commenters agreed with the Committee's recommendations to not endorse 5 of the 6 measures under review with the suggestion to refine the measures to differentiate between facility and provider performance levels.

Developer Response:

The Risk-Standardized PAC Rate (RSPR), which is derived from measuring the base rate of potentially avoidable complications (PACs) for a condition, procedure, or acute event, can be applied to individual physicians, practices, medical groups, facilities and health systems. The determination of the unit of accountability – entity measured – is based on whatever the user of the measure would decide as being appropriate. Methodologically, the predicate to the measurement is a reliability test that determines the minimum sample size required to compare the performance of providers. That sample size requirement will likely drive the decision about the best level of measurement, from the individual physician to the facility. The RSPR does not produce some raw count of individual occurrences of potentially avoidable complications, but rather a risk-standardized rate. This creates an appropriate measure of comparative performance, which can further be stratified as average, above average or below average. Therefore, no one being measured will be penalized for having patients that experience a PAC. Instead, only those that have far higher rates of occurrences than others will have a poorer performance, much like any other composite rate used today.

There is another important consideration about which provider to measure, and that is the attribution of the patient's episode to a provider. There is no standard way of attributing procedures, but there are well-accepted industry conventions, which we have applied in our methods. For example, procedures are often attributed to both facilities and the physician performing the procedure. However, the measure user can make its own determination of attribution.

Therefore, to be clear, potentially avoidable complications are counted within the context of an episode of care. Episodes are then attributed to providers using certain logic. It is the result of that attribution which creates the provider-specific Risk-Standardized PAC Rate. Measure users that want to attribute all procedural episodes solely to facilities, for example, can do that. Others who might want to assign all procedural episodes to a provider group, as opposed to individual physicians can also do that. And the method allows for all of the above, provided the sample sizes are adequate.

Potentially avoidable complications are defined for each episode of care, from a patient-centered perspective. Much of the measurement field today often takes a provider-centric view of measurement, meaning that the starting point is to determine whether the sequelae of a specific intervention in the treatment of a condition is tightly within the control of the physician performing the intervention. HCI3's approach is instead based on whether or not negative sequelae were experienced by the patient irrespective of whether the provider who is attributed the episode perceives those sequelae as being under their control. As recommended long ago by the Institute of Medicine, care should be patient-centered, and the RSPR is therefore designed to be patient-centered. Even if some PACs aren't directly controllable by the managing physician, their occurrence can always be influenced by the selection of high quality upstream and downstream providers.

Proposed Committee Response:

The Committee has reviewed the comments and taken them into consideration prior to the measure reconsideration process for the five HCI3 measures and one measure where consensus was not reached.

Consensus Not Reached Measures

2751: Proportion of Patients Undergoing an Angioplasty Procedure (Percutaneous Coronary Intervention - PCI) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

The Committee did not reach consensus on overall suitability of this measure for NQF endorsement. The Committee agreed that the reliability, validity, and feasibility testing results were adequate. However, the Committee remained concerned that the Type 2 PACs were too broad and that the facility would be held responsible for PACs unrelated to a PCI. The Committee also questioned the equal weighting of PACs, such as the fact that a post-procedural fever was equally weighted with hemopericardium and other serious complications. Although

the Committee agreed with the developer that weighting would be arbitrary, they continued to question the validity of equally weighting of sepsis and fever.

The Committee also expressed concern over the large sample size needed to reach an acceptable reliability level with this measure. Due to the large sample size needed, this measure could not be used to assess low-volume facilities. Lastly, one of the Committee members suggested that reporting PAC rates could potentially lead to unintended consequences such as a reduction in PCIs performed in high risk patients in an effort to reduce the number of PACs.

Developer Response: See developer response above.

Action Item: Committee will revote on overall suitability for endorsement.

Reconsideration Requests

2740: Proportion of Patients with Coronary Artery Disease (CAD) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

2747: Proportion of Patients with Heart Failure (HF) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

2748: Proportion of Patients with Hypertension (HTN) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

2749: Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

2752: Proportion of Patients Undergoing Pacemaker / Defibrillator Implantation (PCMDFR) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)

Developer Rationale for Reconsideration: HCI3 is formally requesting reconsideration of the decision of the Cardiovascular Standing Committee for measures #2740, #2747, #2748, #2749, and #2752. Our reconsideration request is based on the fact that, in our opinion, the Committee failed to vote on the merits of the measures or the specific evaluation criteria, but rather on their acceptability to many members of the Standing Committee as a potential source of "personal" accountability. The evidence provided for most of the submitted measures was and is, for all intents and purposes, identical. The main difference between the first four and the last two was the level of measurement. At several instances members of the Standing Committee expressed support for the

validity of the submitted measures if the level of measurement were a group of providers or health system, but clear antipathy to “being dinged” for complications at the individual level. We believe, none of the evaluation criterion under consideration for a measure ask or suggest that members of a Standing Committee assess a criterion based on personal feelings about how a measure would impact them, and yet the comments made by the Committee members centered almost exclusively on that issue. Further details are provided in the attached Request for Reconsideration letter and Exhibits.

Please see the attached memo and exhibits for more information.

Action Item: Based on comments received and the information provided by the developer the Committee will reconsider these measures for endorsement recommendations.

HEALTH CARE INCENTIVES IMPROVEMENT INSTITUTE®

Fair, Evidence-based Solutions. Real and Lasting Change.

François de Brantes
Executive Director

13 Sugar Street
Newtown, CT 06470
0: 203 270-2906
C: 203-731-4465
Francois.deBrantes@hci3.org

October 23rd 2015

Mary George, MD, MSPH, FACS, FAHA
Senior Medical Officer

Centers for Disease Control and Prevention Division for Heart Disease and Stroke Prevention

Thomas Kottke, MD, MSPH
Medical Director for Population Health
Consulting Cardiologist
HealthPartners

Dear Drs. George and Kottke,

On September 10th 2016 my colleague Dr. Rastogi and I came to present to and answer questions from you and the Standing Committee you chair on several composite outcome measures. The Standing Committee voted on the criteria required to consider the measures for endorsement and the results of the vote are reported in Exhibit 1 attached. To summarize, the Committee voted that four of the six submitted measures were not important to measure, but two were, and that one of the measures sufficiently met the criteria, and one met most but not all.

The comments and discussion during the meeting and leading to the vote on the first four measures displayed, in our opinion, a significant bias by most of the

members of the Committee against measurement at the individual physician level. At several instances members of the Standing Committee expressed support for the validity of the submitted measures if the level of measurement were a group of providers or health system, but clear antipathy to "being dinged" for complications they felt were beyond their control.

There were other specific issues raised by Committee members on the number of diagnosis codes that were included in the definitions of the measures and the relevance of these codes to the particular condition for which the measures have been built. We address these issues more fully in Exhibit 2.

Overall, we are submitting an official request that the Committee reconsider its decision and votes. Our appeal is based on the fact that, in our opinion, the Committee failed to vote on the specific elements expressed in the criteria but rather voted to "kill" our measures based on the above-mentioned bias. In Exhibits 3 and 4 we provide a summary of the evidence in support of our measures on the importance to measure and the other criteria. Mostly, we submit for your consideration that the significant inconsistency with which the Committee voted on the measures should call into question the manner in which the Committee evaluated the merits of the measures.

The evidence provided for most of the submitted measures was and is, for all intents and purposes, identical. The main difference between the first four and the last two was the level of measurement and, prior to voting on the evaluation criteria, several Committee members asked to be reminded about the level of measurement. The last two measures had been tested at the facility level and were therefore under consideration at that level, which seemed to be far more within the comfort zone of the Committee members.

None of the evaluation criterion under consideration for a measure ask or suggest that members of a Standing Committee assess a criterion based on personal feelings about how a measure would impact them, and yet the comments made by the Committee members centered almost exclusively on that issue. Several of these comments included: "I can see myself being held accountable for PACs of type 1, but certainly not for PACs of type 2"; "How can I be held accountable for complications that could occur from patient non-compliance?"; "I'm a surgeon and have no control over what happens post surgery". These comments do not relate, in any way, to any of the evaluation criteria for a submitted measure.

As further evidence of the biases by some Committee members, the criterion for Feasibility of the PCI measure (#2751) received one insufficient vote, one low vote and seven moderate when, from all the evidence submitted, the only objective vote for that criterion was high. As a reminder, the criterion for Feasibility assesses whether a. data are generated and used in care delivery; b. data are electronic; and c. a data collection strategy can be implemented. Given that our measures are calculated using claims data, it's difficult to understand how someone could vote anything but high on that criterion.

While we certainly appreciate the difficulty of assessing the validity of the disparate measures that are submitted for endorsement, the proceedings on September 10th did not, in our opinion, lead to a vote on the merits of our measures or the specific evaluation criteria, but rather on their acceptability to many members of the Standing Committee as a potential source of personal accountability. We therefore formally request that the criteria for measures #2740, #2747, #2748,

#2749, and #2752 be re-evaluated and voted on again.

Sincerely,

Francois de Brantes Executive Director

Health Care Incentives Improvement Institute

Attachments:

Exhibit 1 – results of vote

Exhibit 2 – measure developer response to specific methods questions Exhibit 3 – summary of evidence of measure validity

Exhibit 4 – summary of scientific acceptability

Cc: Leslie Vicale

Helen Burstin

Reva Winkler

EXHIBIT 1 – Standing Committee Voting Results

- **2740: Proportion of Patients with coronary artery disease (CAD) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**
  - Total Votes (17)
    - Health Outcome Rationale --- 3 (18%) yes, 14 (82%) no
      - This measure did not pass importance.
- **2747: Proportion of Patients with Heart Failure (HF) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**
  - Total Votes (17)
    - Health Outcome Rationale--- 2 (12%) yes, 15 (88%) no
      - This measure did not pass importance.
- **2748: Proportion of Patients with Hypertension (HTN) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**
  - Total Votes (17)
    - Health Outcome Rationale--- 3 (18%) yes, 14 (82%) no
      - This measure did not pass importance.
- **2749: Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**
  - Total Votes (17)
    - Health Outcome Rationale--- 5 (29%) yes, 12 (71%) no
      - This measure did not pass importance.

- **2751: Proportion of Patients undergoing an Angioplasty Procedure (Percutaneous Coronary Intervention --- PCI) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**

  - Total Votes (17)

    - Health Outcome Rationale--- 11 (65%) yes, 6 (35%) no
    - Opportunity for Improvement--- 7 (41%) high, 6 (35%) moderate, 2 (12%) low, 2 (12%) insufficient
    - Composite Logic--- 0 high, 11 (65%) moderate, 2 (12%) low, 4 (24%) insufficient
    - Reliability--- 0 high, 11 (65%) moderate, 4 (24%) low, 2 (12%) insufficient
    - Validity--- 0 high, 12 (71%) moderate, 3 (18%) low, 2 (12%) insufficient
    - Composite Analysis--- 2 (12%) high, 9 (53%) moderate, 3 (18%) low, 3 (18%) insufficient
    - Feasibility--- 8 (47%) high, 7 (41%) moderate, 1 (6%) low, 1 (6%) insufficient
    - Usability and Use--- 7 (41%) high, 7 (41%) moderate, 2 (12%) low, 1 (6%) insufficient

    - Overall, the Committee (10 votes, 59%) did not reach consensus on this measure. It will be posted for public comment and the Committee will have the opportunity to revote on the measure during the post-comment conference call.

- **2752: Proportion of Patients undergoing Pacemaker / Defibrillator Implantation (PCMDFR) that have a Potentially Avoidable Complication (during the episode time window) (HCI3)**

  - Total Votes (16)

- Health Outcome Rationale--- (15 votes) 9 (60%) yes, 6 (40%) no

- Opportunity for Improvement--- (15 votes) 2 (13%) high, 11 (73%) moderate, 0 low, 2 (13%) insufficient

- Composite Logic--- 0 high, 8 (50%) moderate, 5 (31%) low, 3 (19%) insufficient

- Reliability--- 0 high, 9 (56%) moderate, 5 (31%) low, 2 (13%) insufficient

- Validity--- (14 votes) 0 high, 5 (36%) moderate, 5 (36%) low, 4 (29%) insufficient

  - This measure did not pass scientific acceptability.

EXHIBIT 2 – Response to Specific Methods Concerns

Significant number of diagnosis codes included in the definition of PACs

Several Committee members expressed concern (some expressed derision, which was not appropriate) about the number of and heterogeneity of the diagnosis codes included in the definitions of Potentially Avoidable Complications. The following presents a brief history of the source of these codes, summary statistics from the materials submitted to the Committee and comparisons with other widely used measures that rely on Diagnosis codes, including measures recently endorsed by the NQF.

A. History of source of codes

Diagnosis codes included in the International Classification of Diseases, revisions 9 and 10 are widely used by all measure developers as a means to define study cohorts, inclusion and exclusion criteria. There are 14,000 ICD---9 diagnosis codes and roughly 70,000 ICD---10 diagnosis codes. While ICD codes are endogenously grouped within the ICD categorization by disease, condition, major body function, codes that can be relevant to a specific population can be included in multiple sections of the ICD classification.

In 1993 the Agency for Healthcare Research and Quality developed the Clinical Classification System as a means to help healthcare services researchers group the 14,000 ICD---9 diagnosis codes into more meaningful groupings. In the researchers' words: "The key factor in creating these categories was the extent to which conditions and procedures could be grouped into relatively homogeneous clusters of interest to public policy researchers."
For a more complete history of the development of the AHRQ CCS, please refer to: https://www.hcup---us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf.

AHRQ's CCS has been used extensively by measure developers over the years, including HCI^3.

B. Codes included in HCI^3's Potentially Avoidable Complications (PACs)

As described in the measure submission materials, PACs are defined as either of Type 1 or Type 2. Type 1 PACs are complications that are relevant to the index condition such as acute exacerbations of that condition or close co---morbidities. Type 2 PACs are complications that are relevant to patient safety failures, whether they occur as a result of a hospitalization related to an acute exacerbation of the index condition or other ambulatory based patient safety failures such as drug---to---drug interactions.

Table 1 summarizes the number of codes and corresponding CCS groups for each of the submitted measures:

Table 1 – Summary of code counts and CCS group counts, by condition

Episode Name	Number of PAC Codes	CCS Category
HTN	811	62
Type 1	294	27
Type 2	517	35
ARRBLK	799	61
Type 1	278	24
Type 2	521	37
HF	846	65
Type 1	310	27
Type 2	536	38
PCMDFR	862	61
Type 1	295	21
Type 2	567	40
CAD	783	52
Type 1	262	18
Type 2	521	34
PCI	868	66
Type 1	328	27
Type 2	540	39

The majority of the diagnosis codes are included in the Type 2 PACs because they represent a far broader range of potentially avoidable complications. For example, PACs of Type 2 include CMS's Hospital Acquired Conditions and AHRQ's Patient Safety Indicators, and of the 500 to 550 Type 2 PAC Dx codes, 140 are HAC codes, and over 200 are PSI codes. HACs and PSIs are broadly recognized as important measures of patient safety and no organization has questioned the underlying number of diagnosis codes that are included in them.

Importantly, and as explained in A, the diagnosis codes included in the PAC definitions come from AHRQ's CCS. Table 1 shows that of the several hundred CCS categories, an average of 25 are defined as PACs of Type 1, and between 35 and 40 are defined as PACs of Type 2.

C. Comparisons with other measures:

As a basis for comparisons, there are dozens of diagnosis codes included in routinely used DRGs such as Cardiac Arrhythmias and Conduction Disorders or Cardiac Defibrillator Implantation. Some of the diagnosis codes included in these DRGs are for syncope and collapse as well as chronic kidney disease with heart

failure. Similarly, the DRGs for Heart Failure and Shock include several dozen diagnosis codes, some of which are unspecific such as an Unspecified shock. The Sepsis DRG has 62 diagnosis codes. In other words, there is significant heterogeneity in the conditions that make---up many DRGs and for which providers are held accountable, especially around readmissions.

In that vein is the NQF endorsed measure for PCI Readmissions (NQF #0695), which was submitted by the American College of Cardiology. The ACC used the same CCS categorization that HCI[3] has used, and includes 100 CCS groups representing several thousand diagnosis codes as valid inclusion criteria for readmissions after PCIs. Some of these diagnosis codes include eye disorders, otitis, and poisoning by non---medical substances. These groups and associated codes are identical to some of the ones included in the HCI 3 definitions of Type 1 and Type 2 PACs, but many many more are not included in the PAC definitions.

In other words, the ACC, in its measure for readmissions following a PCI has included readmissions that have no relationship whatsoever to either broad patient safety issues, the PCI procedure itself, or the underlying conditions that contributed to the need for a PCI. The logic used by the ACC, similar to the logic used by HCI 3, focuses on the importance of the overall management of a patient. And the methods used by the ACC to include these diagnosis codes are identical to that of HCI 3, namely leveraging AHRQ's CCS.

There is a similar NQF endorsed measure on readmissions after CABG surgery (#2514) that was submitted by the STS. Much like the PCI readmission measure, any admission within 30 days after discharge is counted as a readmission. That can include an admission for a severed toe as one of the Committee members derided us about. It's important to note that Potentially Avoidable Complications do not include such a broad definition of complications and certainly not severed toes.

There are close to two dozen other NQF---endorsed measures that include re---admissions post---discharge for any reason other than a planned procedure. Much as described above for the PCI and CABG readmission measures, acute care admissions within 30 days of discharge for reasons that have very little, if anything, to do with the primary reason for the index admission are counted as complications and added to the numerator.

As such, there is a long list of NQF---endorsed measures, including HCI 3's, that take a broad brush in counting what constitutes as potentially avoidable complications. In fact, and to a very large extent, HCI 3's PAC measures take a far more conservative approach to labeling complications than the aforementioned endorsed measures, focusing on complications that are medically related to the underlying condition for which the measure is defined, or well---accepted patient safety measures that are currently being reported.

EXHIBIT 3 – Summary of Evidence on Importance to Measure

Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). 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 (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). 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 (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re---engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found 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.

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.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. 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 (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality 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 validity of its use (Normand 2007; Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001; NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") 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 do occur (Leibson 2008).

References:

1) 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

2) 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.

3) James JT. "A New, Evidence---based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122---128.

4) 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

5) 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.

6) BCBSNC: Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pd

f

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3---5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274---284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing---costs/healthcare---incentives---payment---pilot---hipp/

11) 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.

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

13) 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.

14) 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.

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

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) 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.

EXHIBIT 4 – Summary of Scientific Acceptability

All PAC measures can be consistently calculated by following the measurement criteria. PACs are specific to an episode of care. Episodes of care are clearly defined and all the code sets associated to each defined episode are available as Open Source definitions on the HCI 3 web site (see http://www.hci3.org/content/ecrs---and--- definitions).

Episodes are adjusted for the severity of patients to provide appropriate comparisons. There are two types of risk stratification techniques used in the HCI 3 methodology. The first is a standard list of risk factors that creates a patient's risk profile and is captured from ex---ante or historic claims present before the start of a patient's episode of care. These risk factors can influence the use of services for that patient in the management of the studied condition or procedure. The second is a sub---typology contained within each episode definition in a manner that is analogous to other episodes of care such as DRGs. The DRG sub---typology includes "with Major Co---morbidities" and "with or without Complications". Similarly, the HCI 3 episodes contain sub---types such as CAD with Previous CABG or CAD with unstable Angina.

The use of standard population wide risk factors is widely used in adjusting for patient severity. The most prominent example is Medicare's Hierarchical Condition Categories (see: https://www.cms.gov/Medicare/Health---Plans/MedicareAdvtgSpecRateStats/Risk---Adjustors---Items/Evaluation2011.html?DLPage=1&DLEntries=10&DLSort=0&DLSortDir=desc ending) . There are a considerable number of Condition Categories that contain all 14,000 ICD---9 diagnosis codes. "Although HCCs reflect hierarchies among related disease categories, for unrelated diseases, HCCs accumulate. For example, a male with heart disease, stroke, and cancer has (at least) three separate HCCs coded, and his predicted cost will reflect increments for all three problems." In other words, the HCCs recognize the interaction between conditions/patient risk factors and the models calculate those interactions to determine the effect on the expected costs of care. The number of interactions can be infinite and are only constrained by what is revealed through the analyses of data.

The HCI 3 model operates in a similar way with its list of universal Risk Factors. Each RF contains a number of diagnosis codes related to that particular condition. Each RF is then used as an independent variable to predict the use of services. Any given patient will only have a small subset of all RFs, much like any given patient will only have a small subset of Condition Categories.

Note that during the September 10th Standing Committee meeting, several members of the Committee expressed concern that they would get "dinged" for complications over which they felt they had little control. That statement is

completely inaccurate. The purpose of the severity adjustment model is to create a comparative rate based on the health status of the patients attributed to the individual provider. For example, if a provider's expected PAC rate is 25% and the actual PAC rate is 30%, then the risk---standardized comparative PAC rate score would be 1.2. No provider is "dinged" for the random occurrence of a PAC. It's only when the observed rate is significantly higher than the expected rate that the comparative value will be unfavorable.

To adequately and fairly risk---adjust PAC rates, a logistic regression model is fit to predict the occurrence of a PAC during an episode using the universal risk factors and episode sub---typologies as covariates (Iezzoni 2003). The estimates obtained from the model are used to calculate patient---level probabilities for the occurrence of PACs. The patient---level probability estimates are summed to construct aggregated measures (e.g., facility/provider---level).

To make comparisons between provider practices or facilities, we construct a risk---standardized PAC rate (RSPR). 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 (i.e., mortality, readmissions, etc.). 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 PACs for the provider.
- The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice---level unstandardized performance ratio.
- 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:

$$\text{Adj Outcome}_j = \{(\text{SUM Observed}_{ij})/(\text{SUM Prob}_{ij})\} \times \{(\text{SUM Prob}_i) / (\#\text{ of episodes})\}$$

Where individual *i* is attributed to unit of analysis *j* (e.g., practice, provider, etc.)

Reference used above:

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

This method has proven to generate statistically reliable rates of PACs at the individual physician level for chronic conditions and at the facility level for procedures.

For the reliability analysis, we restricted the data to only providers with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the provider to which they were attributed. 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 the 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.
Reference:
Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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).

Although scores among providers with at least 10 episodes were low, many had scores that met or exceeded the minimum acceptable level for reliability. Moreover, limiting providers to those with at least 25 or 50 episodes, scores were consistently good. These results demonstrate that the measure sufficiently differentiates providers' performance.

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.

Table 2 below summarizes the results of the reliability testing for the submitted measures.

Reliability Scores CAD	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	468 (100)	171 (37)	80 (17)
Median (IQR)	0.73 (0.61,0.83)	0.85 (0.79,0.91)	0.92 (0.88,0.95)
Range	0.50---1.00	0.72---0.99	0.84---0.99
Reliability Scores HF	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	81 (100)	27 (33)	13 (16)
Median (IQR)	0.61 (0.52,0.75)	0.80 (0.75,0.85)	0.85 (0.83,0.87)
Range	0.43---0.94	0.69---0.94	0.80---0.94
Reliability Scores HTN	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	3,702 (100)	2,011 (54)	1,039 (28)
Median (IQR)	0.79 (0.67, 0.89)	0.87 (0.81, 0.92)	0.92 (0.89, 0.95)
Range	0.49---1.00	0.71---1.00	0.83---1.00
Reliability Scores ARRTHM	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	575 (100)	232 (40)	103 (18)
Median (IQR)	0.66 (0.54,0.79)	0.80 (0.74,0.87)	0.88 (0.84,0.93)
Range	0.42---1.00	0.65---0.99	0.79---0.99

A final note on Scientific Acceptability. The NQF criteria for endorsement stipulate that outcome measures have sufficient validity if they have face validity. We have strongly demonstrated the face validity of these measures with the volume of papers that have been written about them and their acceptability to both the research community and, in practice, by payers and providers who use the information as a means to improve the quality and affordability of health care.