

To: David Lansky, Chair, Quality Measurement Workgroup of the HITPC
From: Peter Basch, MD, FACP
Re: Guiding Principles for Meaningful Use CQMs

March 8, 2012

As you are probably aware, after I uncovered issues with reliability and consistency of results of several Stage 1 CQMs, I was asked to consult to ONC regarding future CQM development. As part of that work I conducted an analysis of Stage 1 CQMs; with the specific aim not to find fault, but to find areas of commonality that may have led to definitional problems, which might then inform future CQM development.

The overarching purpose of Meaningful Use is to make it more likely than not that health IT will ultimately be optimally developed, optimally implemented, and optimally used to make care better (quality, safety, value, efficiency, fairness, etc.) – which of course further assumes continuous improvement of technology, workflow, and aligned payment. For providers, it is clear that MU foundational measures are meant to steer the on-boarding process in the right direction. But the “heart and soul” of the program for providers is, and must be tightly connected to, measurable improvement in care.

That said, providers do have a history with quality measurement, and for many, that history is not good. Providers have been exposed to claims-based quality measurement for years, and have come to expect that the reports are wrong more often than right, and thus ignore them. And the fact that claims-based reporting has gotten better doesn't really matter for these providers. Many of them will never trust a “quality report” from a payer again, and more importantly, will not take action on them. However, we have an opportunity with EHR-based quality reports to change that history and from what I have seen with my own work – most providers are willing to give this new approach a chance.

I have also been following recent comments from WG members, specifically those about the purpose of quality measurement in the MU program. While I appreciate the desire to only use CQMs that have been perfected, I believe that the MU program *per se* declares a more pragmatic approach. In my view, it declares that we can start now, and need not wait until EHRs, PHRs, HIE, ePrescribing, etc are perfected. In fact, bringing a large mass of people to the “here and now” both brings enormous market pressure and a vision not typically used by early adopters – such that rapid progress towards “better” is possible.

I have also been struck by the frequency of use of the terms “parsimony” and “exemplars.” While I particularly like the concept of exemplars as expressed by you and Paul Tang, I believe that for the MU program, we should be talking about more than exercising EHRs. At least as it pertains to provider engagement and activation (yes, we need to do this with providers as well as consumers), we must also look to CQMs to exercise care delivery workflow and provider / care team analysis and action plan revision. I would thus strongly disagree with the notion that the results of earlier Stage CQMs are or should be irrelevant – as we are just testing the technology. When engaging a busy provider, and the idea is, “I need a minute of your time....And BTW, none of what you just listened to matters;” we are done and have likely lost a golden moment to change behavior and inspire the desire to achieve excellence with a health IT enabled infrastructure. I would much rather take advantage of provider's drive towards perfection and excellence. We are seeing this now at my organization, where we just

attested for ~200 EPs; and providers are now asking, “how can I do better; what can I do to raise that 81% to 100%.”

I must confess I very much liked the concept that was raised last year at the QM WG meeting of parsimony. I believe I may have added to its rise in stature by saying that embracing parsimony as a guiding principle was a way of avoiding my great fear at that time – measure mania. After living through Stage 1 CQMs, I would like to revise my embrace of the simple and sparse to “strive for parsimony where parsimony makes sense.” Furthermore, if providers are already supposed to be following certain guidelines for particular conditions and categories of patients, and quality measurement should be an extraction or byproduct of providing good care; there may be less value to parsimony than was previously thought.

Below are the Guiding Principles of MU CQMs that I provided to ONC

I. Measures Must Be Important, Understandable and Transparent to Patients and Providers

Look to the foundational measure definitions as a guide. One should not need to hire a dB engineer or quality measurement expert to understand a measure. It should be clear on its face to patients and doctors. Why is this measure here? What is the measure referring to? Which patients are being measured (denominator)? What is being measured and during what period of time and what counts as “satisfied the measure” (numerator)? Are there exclusions, and if so, how can one be claimed? Lastly, when something is highly relevant, clear and transparent, is it also clear as to what actions can be taken to improve the measure? And as these measures are for EPs and EAs, can the actions of the EP or EA improve a measure score?

II. Measures Must Reflect Current Evidence-Based Medicine

The fact that a measure is understandable doesn’t mean that it is right. As an example, the existing Stage 1 diabetes CQMs include a denominator of a patient with diabetes on the problem list and a visit with the provider during the measure period; OR metformin on the active medication list without a diagnosis of polycystic ovaries. While one can surmise that the measure authors were attempting to capture a proxy for diabetes (when it wasn’t listed on the problem list but most certainly existed), they of course could not have foreseen that more recent guidelines would call for use of metformin for prediabetics. The unfortunate result is that doctors who are most up to date with ADA guidelines for treating prediabetics now have the worst diabetes quality scores – as least as determined by Meaningful Use Stage 1 CQMs.

III. Measures Should Use Consistent Terminology and Definitions

Whereas a measure author or steward may have addressed a condition using a particular definition of set of conditions for inclusionary characteristics (and thus meet guiding principle #1 above); if we expect patients, providers, and vendors to actually understand, operate upon, and/or incorporate those measures into their respective realities, inclusionary and exclusionary characteristics should be consistent, wherever possible. For example, if there are five quality measures for diabetes, and the title of the quality measure does not specifically call out a particular feature – such as “Medical Management of Diabetics with Renal Failure;” it would only lead to confusion and poor adherence if one measure defined characteristics of inclusion differently (for example), solely because of a difference in authorship.

IV. Measure Denominators Must Consist of Patients Seen By the Provider During the Measure Period

One essential element to achieving provider buy-in is ONLY measuring patients seen during the measure period, unless there is a clear, compelling and reasonable alternate denominator definition. Returning again to the examples of the Stage 1 CQMs for diabetes, the denominator definition of EITHER a diagnosis and at least one visit OR an active medication (in that case, metformin), created even worse havoc with denominators. First, new users of EHRs who followed recommended best practices of preloading key patient data (such as problem list, medication list, allergy list, etc.), found that these preloaded patients who were not seen during the measure period and thus may have been missing key clinical data, were nevertheless counted in the denominator. Veteran EHR users had a similar problem with patients who left the area or died. In EHRs when a patient status changes (such as deceased, inactive), that status is reflected by a global change to the patient (deceased, inactive), and not by removing medications from a medication list. Thus, many veteran EHR users had patients who hadn't been seen in many years or were deceased, counted in the diabetes CQM denominators. Where the denominator does not ONLY measure patients seen during the measure period, that distinction must be clearly called out.

V. Measures that are Meant to Have Broad Applicability (such as Core Measures) Must Include Numerators That Reflect Reasonable Actions Taken By Providers in All Applicable Specialties

Key measures, such as core measures must include reasonable actions that are consistent with good medical practice and the scope of practice for the provider being measured. For example, if a measure for influenza vaccination only counts vaccine administration, the measure would either provide an inaccurate reflection of current practice, or lead to providers pressuring patients to stop doing what is commonly done (vaccines obtained where convenient and inexpensive – such as at work, at church, at a drugstore, etc).

And as per scope of practice, an internist, endocrinologist, family physician, or pediatrician may be comfortable with counseling for BMI out of range. A specialist may appropriately feel that it is outside the scope of his/her practice to provide such counseling; in which case, good medical practice would dictate that documentation of “referred to PCP for weight management counseling” is appropriate. Many specialists already believe that Meaningful Use and EHRs were primarily designed for primary care, and to have broadly applicable measures that don't respect the scope of their practice would add to that concern.

Homeland Security advertises an action plan for the American people – and this action plan does not require that ordinary citizens investigate suspicious packages, or attempt to apprehend suspicious persons. Rather, it is advised, “if you see something, say something.” I believe this is a good practice for the country, and a good approach to emulate in the Meaningful Use program, as it allows all providers to act on key or core measures within the scope of their practice, and in the common good of their patients. As an example, because of Meaningful Use, many specialists are now capturing vital signs. While we don't want a provider who is not competent to treat hypertension to just feel as if they are data collectors for others; we could – if the CQM numerator was appropriately broad – also enlist them in the campaign to use health IT to improve care. Thus, any provider who takes a blood pressure could be rewarded for “seeing something, and then saying something;” where the best action for that provider might be recognizing that the blood pressure may be out of range, and further recommending follow-up with the PCP or principle provider.

VI. Wherever Possible, Measures Should be Chosen or Constructed Based On All of the Above AND Their Relevance to Key Health and Healthcare Aims

For Stage 1, many of the CQMs appear to have been chosen because they already existed and were close to what was needed – but in fact as a less than ideal fit. The problem with these “in the ballpark measures” is that providers are trained to look for exceptions, and measures that are close but have something of significance that is “off” will continually remind providers that they are striving for a measure they know to be wrong. And if the goal is to move from reluctant compliance to full buy-in; this could be a serious problem. For example, if ONC and CMS wanted to have CQMs to support the Million Hearts Campaign, one could look to existing CQMs measures and see if there was one for aspirin use, one for blood pressure, one for lipid control, and one for cigarette smoking. However, unless those measures were fully consistent with the published literature and national campaign goals, providers will see them as nonsensical and not as a pathway to clinical excellence. In such cases, it would be better to send the measure back to the measure author for updating.

As an example, the clarification to the adult weight management measure (as published by ONC to the RECs) specifies actions that may have been appropriate at some point in time, but certainly not in 2012. Numerator activities include vitamin injections, lap band and gastric bypass surgery, and formal nutritional counseling. I am not sure why a CPT code for vitamin injections is still listed as a numerator activity, but it certainly shouldn't be. Also, if this measure is meant to address the growing obesity epidemic, instead of using this measure that “sort of fits,” it should either address all patients with a BMI out of range, and start with a description of numerator activities that are reasonable, evidence-based, and appropriate for the vast majority of patients who will fall into the denominator (depending on the patient population, more likely to see BMIs 25-30, or 25-35) – and lap band and gastric bypass should not be the first numerator activities on the list. Or, a measure could be constructed to address the most serious outliers, where more extreme actions are more typically advised.

However, as currently constructed, the numerator does not appear to allow for judgment or prioritization. A fit individual with a BMI of 25.01 seems to require the same numerator actions as a person with a BMI of 50. A measure such as this should be sent back with a focus on, what might be a reasonable interpretation and action for patients who have a BMI > 25, but are judged not to need weight loss; for patients with a BMI > 25 and are mildly overweight vs. obese; for patients with a BMI > 25, but are already committed to a weight loss program, etc. If we cannot have a measure that allows for sound medical judgment to be judiciously applied (or not if not appropriate), then we risk losing the battle for the “hearts and minds” of providers in this campaign against obesity; and risk a similar fate with patients.

VII. Measures Should Be Chosen in the Context of Other Measures

If a core or menu set foundational measure addresses a condition or describes actions for a condition, and a CQM addresses the same condition and/or actions, wherever possible these should be harmonized. For example, there is a core foundational measure for smoking status. What would make sense is for the QM WG to request of NQF a companion CQM to use for “current smokers” – as identified by use of the MUM. Instead, we have a completely separate measure pair for tobacco status and counseling, which is confusing to patients, providers, and vendors. If it takes a health IT expert or a measure guru to explain to providers the subtle differences between measures – we risk turning providers who currently support a national effort at smoking cessation to be become antagonistic, or worse, apathetic.

Another example... There is a MUM for vital signs that is clearly defined. And while the MUM threshold for Stage 1 is 50%+; the message to providers is that this is a minimum threshold and that the desired behavior is current VS (including BP on all patients > 2 years of age). The CQM for BP is similar, and offers a distinction without a difference. Again, if we want providers to embrace this use of EHRs as a campaign to advance health and healthcare, examples such as these will lead to providers rolling their eyes to the ceiling and dismissing the effort as “one more stupid government program.”

VIII. Measures Should Not Introduce Competing Requirements to Other Measures

The MUM for vital signs is not clear regarding when VS must be taken, but a reasonable assumption is new VS (at least BP and weight) at least annually. The companion CQM for weight management references BMI within 6 months – without specifying that 6 months is +/- 6 months of the visit, or +/- 6 months of the measure period. We could end up with providers enthusiastically (or reluctantly) becoming “good citizens” in this campaign against obesity – and then “reward” these good citizens with “gotcha’s” on the CQM (“thank you for getting a BMI and helping to guide your patient towards an ideal weight. However because you followed instructions for the MUM and not the CQM, your quality score is zero; have a nice day”). This is obviously not acceptable. Providers who buy-in are saying, “What exactly do I need to do, and I will do it.” This is of course also necessary for EHR vendors regarding reporting definitions and companion CDS support.

IX. Specified Data Elements Necessary to Determine a Measure Denominator, Numerator, and/or Exclusion must be “EHR Capable”

Measures are not necessarily constructed by measure authors who are familiar with EHR capabilities, and where this occurs, a measure might presume a finding or calculation that is not typically found in an EHR, or even findable in an EHR. There are 4 subcategories of EHR capable:

- a. Those capabilities that are currently required by certification
- b. Those capabilities that are commonly available
- c. Those capabilities that are seen only by innovators at the “leading edge”
- d. Those capabilities that may fit subcategories a, b, or c – but may be a less than ideal fit because the ability to populate this data element is too variable to be deemed reliable.

For the purposes of CQM development, a measure should be based on required functionality. A measure could declare an intent in future iterations for more advanced requirements, as suggested by current work of innovators.

X. Measures Should Be “EHR Enabled”

Older measures are sometimes based on paper-based capabilities and ignore what is currently or easily capture able within EHRs. This lack of awareness and vision is limiting, and should be addressed by measure authors and stewards, such that future measures take advantage of existing and emerging EHR capabilities.

XI. Measures Should Be Considered as Organic and Part of a Continuous Learning Process

Many / most existing measures were developed for use by paper chart abstractors (who could apply judgment in declaring a denominator, numerator, and/or exclusion) and were then “e-Spec’ed;” and as such do not have a track record for reliability and usefulness based on large numbers of providers using EHRs. We should not assume that what seemed to work in a particular circumstance still works well in an EHR environment where measures are automatically abstracted; or that even if acceptable, a measure could not be made better based on its being used within an EHR. Furthermore, even where measures have been used extensively by many providers using EHRs, we cannot assume they are where they should be. Consider the previous and current messaging surrounding PQRS measures.

Providers have been encouraged to report data, and paid for that reporting. There has been no ask of providers to QA measures. In some circumstances providers have never looked at their scores, and have been satisfied to have any data in the required reporting fields. In other cases, providers may have looked at scores, and if the number looked ok, felt reasonably satisfied. However, we have not approached the provider community with the following statements... “In this new world where performance and quality are everybody’s business, you should carefully choose measure that are within the scope of your practice; and then test these measures by looking at measure exception reporting. Are the patients listed in the denominator patients you have seen during the measure period? Do these patients have the inclusionary characteristics of the measure? Was the numerator activity performed but now showing in the report?”

The rigor of this approach would resolve many of the problems we see today with seemingly good quality measures, and would detect and help lead to resolution of problems with vague definitions, poor programming, results or documentation not being detected for a variety of reasons, etc. In addition to “exercising” the measures and the EHR, it creates a strong ownership by providers. We are essentially saying to providers, during this period of time, you are verifying that what will ultimately become your public face of quality, safety, and value (to patients, to payers, to colleagues) is appropriate and accurate. To use the current mantra for Stage 2 MU, it is one activity that provider should engage with to make meaningful use meaningful.