Developing a Framework and Research Agenda for Overuse and Appropriateness Measures

Agency for Healthcare Research and Quality (AHRQ)
Appropriateness Small Conference
June 9, 2009

Grant Number: 1R13HS018105-01

The Physician Consortium for Performance Improvement®
Convened by the American Medical Association

This conference was supported by the Agency for Healthcare Research and Quality and the National Quality Forum
ATTENDEES

Margaret E. O’Kane
National Committee for Quality Assurance
okane@ncqa.org
202-955-5100

Helen Burstin
National Quality Forum
hburstin@qualityforum.org
202-783-1300

L. Gregory Pawlson
National Committee for Quality Assurance
pawlson@ncqa.org
202-955-5170

John Cooper
Centers for Medicare & Medicaid Services
john.cooper@cms.hhs.gov
410-786-0524

Phyllis Torda
National Committee for Quality Assurance
torda@ncqa.org
202-955-5180

Joyce Dubow
AARP
jdubow@aarp.org
202-434-3901

Karen Adams
National Quality Forum
kadams@qualityforum.org
202-783-1300 x147

R. Adams Dudley
University of California, San Francisco
adams.dudley@ucsf.edu
415-476-8617

Joseph Allen
American College of Cardiology
jallen@acc.org
202-375-6463

Fred Edwards
American Society of Thoracic Surgeons
fred.edwards@jax.ufl.edu
904-655-0427

David Atkins
Department of Veterans Affairs
david.atkins@va.gov
202-461-1517

Lawrence Fine
National Heart, Lung and Blood Institute
finel@nhlbi.nih.gov
301-435-0305

David Baker
Northwestern University
dwbaker@northwestern.edu
312-695-8630

Terry Gilliland
Kaiser Permanente
Terry.M.Gilliland@kp.org
301-272-4913

Mary Barton
Agency for Healthcare Research and Quality
Mary.Barton@AHRQ.hhs.gov
301-427-1638

Robert Hendel
American College of Cardiology
rhendel@midwestheart.com
630-310-2162

Joel Brill
American Gastroenterological Association
Joel.brill@gmail.com
602-418-8744

Peter Hussey
RAND Corp.
Peter_Hussey@rand.org
703-413-1100
George Isham
Health Partners
george.j.isham@healthpartners.com
612-883-6769

Jeffrey Kelman
Centers for Medicare & Medicaid Services
jeffrey.kelman@cms.hhs.gov
202-690-6319

Karen Kmetik
AMA-Physician Consortium for Performance Improvement
karen.kmetik@ama-assn.org
312-464-4221

David Lansky
Pacific Business Group on Health
dlansky@pbgh.org
415-281-8660

Grace Lin
University of California, San Francisco
grace.lin@ucsf.edu
415-476-8617

Susan Milner
National Committee for Quality Assurance
milner@ncqa.org
202-955-1736

Peter Neumann
Tufts New England Medical Center
pneumann@tuftsmedicalcenter.org
617-636-2335

Mai Pham
Center for Studying Health System Change
mpham@hschange.org
202-554-7571

Meredith Rosenthal
Harvard University
mrosenth@hsph.harvard.edu
617-432-3418

Joachim Roski
Brookings Institution
jroski@brookings.edu
202-797-6281

Bernie Rosof
North Shore-Long Island Jewish Health System
svpmd@aol.com
516-465-8260

Dick Salmon
Cigna
Dick.Salmon@cigna.com
860-226-2906

Sarah Scholle
National Committee for Quality Assurance
scholle@ncqa.org
202-955-1726

Sandy Schwartz
University of Pennsylvania
schwartz@wharton.upenn.edu
215-898-3563

David Seidenwurm
Radiological Associates
seidenwurmd@radiological.com
916-784-2277

Tom Valuck
Centers for Medicare & Medicaid Services
Thomas.valuck@cms.hhs.gov
410-786-7479

Michael van Duren
Sutter Health
VanDurM@sutterhealth.org
916-854-6936

Nancy Wilson
Agency for Healthcare Research and Quality
Nancy.Wilson@cms.hhs.gov
301-427-1310
AMA-PHYSICIAN CONSORTIUM FOR PERFORMANCE STAFF

Beth Tapper  
AMA-PCPI  
Beth.Tapper@ama-assn.org  
312-464-4956  

Greg Wozniak  
AMA-PCPI  
Greg.Wozniak@ama-assn.org  
312-464-4956

AMERICAN MEDICAL ASSOCIATION STAFF

Jennifer L. Shevchek  
AMA  
Jennifer.Shevchek@ama-assn.org  
202-659-0599

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY STAFF

Elise Berliner  
AHRQ  
Elise.Berliner@AHRQ.hhs.gov  
301-427-1612  

Steve Phurrough  
AHRQ  
Steven.Phurrough@AHRQ.hhs.gov  
301-427-1617  

Bill Encinosa  
AHRQ  
William.Encinosa@AHRQ.hhs.gov  
301-427-1437  

Kim Wittenberg  
AHRQ  
Kim.Wittenberg@AHRQ.hhs.gov  
301-427-1488

Kenneth Lin  
AHRQ  
Kenneth.Lin@AHRQ.hhs.gov  
301-427-1888
CENTERS FOR MEDICARE & MEDICAID SERVICES STAFF

Susan Ardai
CMS
susan.arday@cms.hhs.gov
410-786-3141

Rosemarie Hakim
CMS
Rosemarie.hakim@cms.hhs.gov
410-786-3934

Mary Kapp
CMS
Mary.Kapp@cms.hhs.gov
410-786-0360

NATIONAL COMMITTEE FOR QUALITY ASSURANCE STAFF

Brooke Barrash
NCQA
barrash@ncqa.org
202-955-3597

Aisha Pittman
NCQA
pittman@ncqa.org
202-955-5162

Robert Saunders
NCQA
saunders@ncqa.org
202-955-1746
Meeting Summary

CREATING A FRAMEWORK FOR MEASURING APPROPRIATENESS

Where does cost fit into the appropriateness framework?

- Include cost in the framework.
  - Cost should be included, implicitly or explicitly. The American College of Cardiology (ACC) originally followed the RAND model, which excluded cost, but added it to the methodology later.
  - Separating clinical effectiveness and cost effectiveness may be best politically, but it is better to address these concepts holistically.
  - There are accepted methods for assessing cost effectiveness. NCQA should take the lead in using these methods.

- Exclude cost from the framework.
  - There is a benefit to disaggregating clinical effectiveness and cost effectiveness. Cost should remain separate and be integrated at the end.
  - Willingness to bear cost is not the same across all populations.
  - At high levels of effectiveness, is cost an issue? Cost becomes more important as benefits become fewer.
  - If we include cost, we must address its variability and who bears the cost.

*Use variation and clear overuse in lieu of predicted effectiveness*

- Misuse:
  - Do not eliminate misuse or attempt to easily define it.
  - Using a standard of harm for misuse sets the bar too high.

- Variation:
  - It is difficult to predict effectiveness; begin with variation, as demonstrated by the Dartmouth Atlas of Health Care.
  - Prioritize areas for measurement based on variability.
  - In the RAND method, most care is uncertain (i.e. neither clearly appropriate or inappropriate). There is pervasive uncertainty in clinical effectiveness.
  - In accountable care organizations, opportunities for improvement are to reduce variation and sensitize clinicians about bias.
  - Seeing the variation data changes physician behavior immediately.
    - Change physician behavior by aggregating data across the community and making data available.
    - Variation data are powerful—relative frequency by severity adjusted condition.
    - Variation is not directly associated with inappropriate care. We must look at the rate of negative test results. If the rate of positive test results is 5 percent, testing is too frequent.

- Overuse:
  - There are different types of overuse: clinically harmful overuse; overuse that is not cost effective; overuse of appropriate care—separating these would help.
– 80 percent of care is uncertain but 10 to 20 percent of care can be defined as some form of overuse. It is not possible to get to 0 percent, but reducing overuse to 5 percent will have a big effect.
– The framework for overuse should not be any different than the framework for underuse.
  • As with underuse, can a simple test work: if the rate is higher or lower than the community’s or the national average than it represents a lack of appropriate care?
– Input into decision making is the same, even though the patient and physician may weigh them differently—symmetry between overuse and underuse is a good thing.

**The patient’s perspective**

• Understanding “cost” to various players gives a clearer picture.

• Since most care is uncertain, the patient’s perspective is highly important:
  – One of the uses of predicted clinical utility is benefit design, but this primarily considers only the doctor’s point of view.
  – The patient’s perceptions of the benefits and risks may alter to rating of care from appropriate to inappropriate.
  – The patient’s perception is that the care is uniformly necessary. We must work to change this perception.
  – Many coverage decisions are arbitrary and based on someone else’s values.
  – The public does not fully grasp the role of opinion and values in decision making.
  – We must have public buy-in to be successful. We must start with what we are really sure about, and be transparent about it.

• How often do we allow an exception for patient preference?
  – There is distinct variation in provider practice with regard to physician and patient characteristics:
    • Some variation is related to physician incentives.
    • Overuse is reduced when physician incentives are based on quality (e.g., acute low back pain measures).

• Doctors’ response:
  – How should physicians address patients who expect care that is not needed?
  – Much of care is influenced by local standards. While it may be against guideline recommendations for a given procedure, it may be the local standard. This introduces concern about malpractice.
  – There is improvement as soon as physicians know performance is being tracked.
Measurement strategy

- Identifying a small number of indications can account for a lot of overuse and will dramatically change provider practice.

- Measures must use all categories of data:
  - Claims based
  - Clinically enriched
  - Chart review

- Create an overarching methodological approach for overuse, underuse and misuse.

- Look at indications rather than at tests. Multiple specialties compete for the same patient; it may be hard to determine who is accountable when we only look at a test—is the ordering physician accountable, or the conducting physician?

OPPORTUNITIES FOR IMPLEMENTATION AND USE OF MEASURES OF APPROPRIATENESS AND OVERUSE

What can we do now?

- Support shared decision making:
  - Use existing patient surveys to determine if patients are informed, assess the level of shared decision making.
  - Determine true outcomes. For example, CABG is often measured in increased blood flow but does not consider patient outcomes, such as cognitive impairment.
  - Create national standards for criteria in the development of measures. Involve patients in measures development.
  - Guide patients in shared decision making. Most of this falls on the primary physician:
    - Patients often do not have the medical literacy to be involved in shared decision making.
    - Patients frequently want inappropriate care (e.g., antibiotics for a URI).
  - Ensure that physicians have information about other physicians to whom they refer patients (e.g., whether a referred physician scores high or low on any given metric).
  - Organize the care system. Shared decision making is not a reality in an unorganized system.

- Incorporate measurement into existing programs:
  - Incorporate key elements for better assessment of appropriateness. The Centers of Excellence (CoE) criteria do not focus on appropriateness.
  - Some CoE institutions demonstrate high utilization, knowing a group’s risk-adjusted utilization rate provides significant information.
• Measurement strategies:
  – Start with “low-hanging fruit” (D-recommendations and NPP report areas).
  – Start with a tough issue that has high variation, where everyone agrees that there is an issue but there is no clear solution.
• Measurement processes:
  • For preference-sensitive care, where a patient enters the system frequently dictates the care he or she receives: options may not be presented to the patient depending on where he or she entered the system.
  – Measuring processes may be useful; going beyond informed consent, looking at where options are presented.
  – It is easy to “cheat” when measuring processes. It is preferable to assess outcomes and patient knowledge.
• Gall bladder removals are an example of preference-sensitive care for the physician: the risk of laparoscopic surgery is minimal, so the community standard is to perform the procedure laparascopically.
  – We need comparative-effectiveness research in this area.

• Put payment policies into place.
  – Change the payment policy to support shared decision making for vulnerable populations.
  – Create fundamental payment reform to reduce inappropriate care. Without bundled, value-based purchasing, evidence and shared decision making will not get far.
  – If physicians benefit from providing certain procedures, their decisions about the procedures can be affected.
CRITERIA FOR PRIORITIZATION

- Rates of appropriateness are same in high and low utilization areas; we must address the gray areas.
  - Labeling gray areas as black or white is dangerous.
  - Elucidate the subtlety in decision making that lies below the surface of gray areas. This will require using more than administrative data.
  - Future research should focus on these grey areas.

- Shared-decision making:
  - Cannot be generalized across all procedures.
    - Will not drive down the use of low-risk procedures.
    - Only works when patients can be informed about actual risks.
    - May not affect grey areas, as it is unclear as to how much shared decision making may affect variation.
  - Must address multiple providers.
    - Primary care physicians reside in the community and rarely visit the hospital, as hospitalists now provide inpatient care for a primary care provider’s patients. The result is a loss of “unspoken” oversight, which leads to overuse (e.g., additional colonoscopies, anesthesiologists sedating average-risk patients for endoscopies).
    - Use a multispecialty approach to shared decision making—community decisions made across providers.
    - QI efforts should support primary care physicians working with specialists to review cases.

- Prior authorization and structural measures:
  - Are there “low-hanging fruit” structural measures that can be used?
    - Ownership of testing facilities, for example, drives a lot of explained variation.
  - Preauthorization is reemerging as decisions are based on overall spending.
    - Preauthorization decreases all care, not just inappropriate care.
    - Identify three or four areas of “low hanging fruit” for measure development.
    - Plans would prefer an alternative to preauthorization. Overuse measures can provide an alternative and reduce preauthorization.

- No single measure will address overuse; we must look at a portfolio of measures.
  - Understand and review a mix of data. Cross-sectional and longitudinal data show that it is possible to be a high performer in some areas and a very low performer in others.
  - Define a minimum data set for appropriateness for a particular condition (e.g., cardiac imaging).
  - Set up a mechanism to study what is being done, while it is being done.
PRIORITIES FOR MEASURES DEVELOPMENT

Set priorities

- Target areas that can be linked to outcomes.
  - Focus on indications, not procedures, as focusing on a given procedure may have the unintended consequence of increasing other procedures.
  - Link to reimbursement; align incentives to reward increased efficiency.
  - Create a feedback loop, looking at false positives and repetitive testing.
    - Look at the number of false positives vs. true positives: a high number of false positives is probably a result of screening the wrong people.
      - In some populations, there is more harm in the aggregate from too many positive results (e.g., mammography).
      - Discern validity (specifically for those of intermediate specificity) of certain diagnostic and screening tests.
  - Repetitive testing is a major issue with diagnostic tests. We need better decision support tools.
  - Collecting additional information using decision support tools can inform research and guidelines about effectiveness and help reduce unnecessary testing.

- Target areas where variation, spending and rate of increase are high.
  - Identify areas where overuse is a bigger problem than underuse.

- Target areas where administrative data can be used.
  - For some issues determining risk is easy clinically but difficult to capture in administrative data.
  - Targeting administrative data could result in favoring uncoded procedures over coded procedures.
  - Target areas that can be generalized for a larger population and think of additional data that would be needed to determine appropriateness.

- Address a difficult topic area to set the standard.
  - Combine outcome measures with appropriateness data.
  - Link clinical databases with administrative databases to assess long-term outcomes. If there is no difference, care was inappropriate.

- Create new criteria to prevent gaming.

- Send a directive to medical societies, suggesting specific measures.
  - Adopt other organizations’ metrics and previous work as a starting point.

**Attribution: how similar is it to underuse?**

- In underuse, the physician who performed the service is the person tied most closely to it.

- In overuse, there is increased sensitivity because of the revenue produced.
  - Procedures should be removed if accountability cannot be determined.
• There is the issue of small numbers with many measures. Measurement will need to take place at the health plan level or at the county or regional level.
  – Reporting for physicians or physician groups will be difficult. If we look at a portfolio of measures, everyone will be “normal.”
  – Explore all loci of accountability: plan level, groups, accountable care organizations.

RESEARCH NEEDED FOR TESTING AND IMPLEMENTATION

The role of panels

• Discrepancies between panels are problematic. Explicit rules and built-in assumptions lead to greater concordance among panels.

• Align appropriateness criteria with guidelines.

• In the RAND approach, the criteria would have been more successful if there was discrimination between 6 and 7.
  – There is a discrimination/calibration issue regarding where to draw the threshold.
  – Look at the clinical and economic costs for these areas.
  – The panel should focus on areas of disagreement and uncertainty.

• Limit the emphasis on panels because it is a step backward.
  – Emphasize data.
  – Do not discount the value of panels because they can be used to synthesize evidence.

Develop feedback loops

• We have underestimated the need for better methods of comparative effectiveness research and real-world pragmatic trial data.

• Interactive feedback loops capture needed data.
  – The trend is toward increased use of diagnostics. We need information on clinical effectiveness.
  – Using appropriateness as an optimization exercise has occurred in many areas, but not in health care. Such efforts do not need to be too complex.

• Discern what is driven by patient preference.
  – Current payment models rely on patient satisfaction. Need to determine physician behavior under other payment models.
  – How do we move patients toward preferring what is necessary?
Where do we want to be?

- Establish research designs.
  - There are many factors in practice, so it is difficult to test.
  - Natural experiments would look at preauthorization, disability, SES.
  - Build on top of the cross-institutional feedback loops that the American Medical Group Association uses.

- Determine data needed for EHR and clinical decision support.
  - Identify the most important decision points in a measurement decision tree, then work with guideline developers to address those areas.
  - Use data generated from feedback loops to determine what data are needed.
  - Determine implementation opportunities.
    - Standards for EHRs are being defined now; we must map activities to measure development.

- Create a standardized model for accountability/attribution (e.g., Is the ordering physician accountable? The specialist? The medical home?).
  - Most measures are not done at the physician level or the physician group level.
  - Report at the group level but provide physician-identifiable data for internal use.