AMERICAN COLLEGE of CARDIOLOGY FOUNDATION
2400 N Street, NW
Washington, DC 20037-1153
(202) 375-6000
(800) 253-4636
Fax: (202) 375-7000

American Heart Association.
Learn and Live.
National Center
7272 Greenville Avenue
Dallas, Texas 75231-4596
Tel: 214-373-6300
Fax: 214-373-9818

June 17, 2010

Helen Burstin, MD, MPH
Senior Vice President for Performance Measures
National Quality Forum
601 Thirteenth St, NW
Suite 500 North
Washington, DC 20005

Via e-mail: performancemeasures@qualityforum.org; hburstin@qualityforum.org

Dear Dr. Burstin:

Thank you for the opportunity to comment on the draft report of the NQF Evidence Task Force: *Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement.* The American College of Cardiology Foundation (ACCF) and the American Heart Association (AHA), support your efforts to provide more explicit guidance for steering committees and technical advisory panels to consistently apply the NQF measure evaluation criteria across projects. This is a very thoughtful report and we especially applaud the emphasis on quantity, quality and consistency of evidence. We support many of the principles outlined in the report, however, we do have concerns related to some of the recommendations and respectfully offer the following comments for your consideration.

The ACCF and the AHA have for many years taken a leadership role in promoting high-quality, evidence-based, patient-centered care for cardiovascular disease, including the development of clinical practice guidelines and performance measures in high priority areas. We have jointly engaged in the production of clinical practice guidelines for over 25 years. Building upon our years of experience in this area, we are continuously refining our processes to ensure transparency and rigor and--most importantly--that adherence to our guidelines contributes to improved patient outcomes.* Reporting on the quality of the evidence, the treatment gap, and the opportunity for improvement, as recommended in the draft report, are laudable goals and fully aligned with the goals of the ACCF and the AHA. Recommending improvements in the evaluation process of the database is also commendable and is a major focus of our joint ACCF/AHA Task Force on Practice Guidelines. However, we are very concerned that the recommendations related to systems for rating of evidence are too prescriptive. After carefully reviewing the draft report, the basis for the stated preference/requirement for USPSTF or GRADE remains unclear to us.

We would strongly urge that the NQF recommend a set of principles/criteria rather than identifying a specific “preferred” or required methodology for grading evidence, especially

in view of the pending Institute of Medicine (IOM) report. The IOM report is expected to recommend best practices for the creation of scientifically valid and trustworthy clinical guidelines, including evaluation of evidence, and protocols to ensure that they are unbiased. The ACCF/AHA has contributed to and supports this comprehensive, ongoing effort by the IOM and looks forward to their recommendations.

Our other major area of concern relates to the apparent requirement that measure developers undertake an independent review of the primary evidence (lines 131-139 in the draft report), a task already performed by guideline development committees. The report does seem to give mixed messages on this point and, if this is not the NQF's intention, this should be clarified. We believe that such a requirement would not only be an inefficient use of time and resources, since it duplicates the work of guideline development committees, but would substantially increase the burden on measure developers. It is consequently likely to slow measure development, which is certainly not the NQF's goal in an era when more and better measures are urgently needed. Our joint ACCF/AHA Task Force on Practice Guidelines is phasing in new methods for guideline writing committees to better present the specific evidence upon which their recommendations are based, and provides explicit information in all guideline documents regarding the scoring system used to rate the evidence. We have also been working over the past several years to create systemic links and to improve communication between guideline and measure development groups. These efforts are best practices derived from our extensive experience developing both guidelines and performance measures and should obviate the need for an additional, independent assessment of the evidence by measure developers.

It is worth noting that some of our reviewers interpreted the draft report as setting the bar for evidence too low. For example, some reviewers questioned whether a single small RCT or even a well-designed observational study would be adequate to pass criterion 1c. Upon carefully reviewing Tables 4 and 5, this does not appear to be the intent, however, it might eliminate some of this confusion if specific examples were provided to demonstrate how a steering committee or TAP would actually go through the process of evaluating a candidate measure using the guidance in the draft report.

The report states that expert opinion is acceptable evidence, that it should be systematically assessed and fully described, and that it will be evaluated as outlined in Table 4 (lines 366-367). It is still somewhat unclear how even systematically assessed expert opinion would be evaluated, on *quantity of evidence* in particular, and would pass criterion 1c. We would like to see this discussion further developed in the report, especially given that RCTs are generally not possible in certain areas (e.g., diagnostic imaging) where there is also an intense demand for more and better information on quality, cost and appropriate use.

Thank you for your consideration of the comments above. We support the NQF's efforts to improve the rigor and promote the consistency of the measure evaluation process and would be happy to discuss our concerns with you directly at any time.

Very truly yours,

Clyde W. Yancy, MD, FACC, FAHA, FACP
President, American Heart Association

Ralph W. Brindis, MD, MPH, FACC
President, American College of Cardiology

cc: Frederick Masoudi, MD, MSPH, FACC, FAHA
Alice Jacobs, MD, FACC, FAHA
Joseph Drozda, MD, FACC
Ray Gibbons, MD, FACC, FAHA
Jack Lewin, MD
Janet Wright, MD, FACC
Rose Marie Robertson, MD, FAHA
Meighan Girgus
Gayle Whitman, PhD, RN, FAHA, FAAN
Charlene May, ACC Staff
Penelope Solis, AHA Staff

* Peterson ED, Roe MT, Mulgund J, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. JAMA. 4-26-2006;295:1912-20.

Boehringer Ingelheim

Karen Pace
Senior Program Director
National Quality Forum
601 13th Street NW, Suite 500 North
Washington DC 20005

June 18, 2010

RE: Review of Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement in the Report "Importance to Measure and Report"

Hemal Shah, PharmD
Executive Director
Health Economics & Outcomes Research
Boehringer Ingelheim Pharmaceuticals, Inc.
900 Ridgebury Rd/P.O. Box 368
Ridgefield, CT 06877-0368
Telephone (203) 798-9988

On behalf of Boehringer Ingelheim Pharmaceuticals, Inc. (BIPI), I am pleased to provide comments in response to the National Quality Forum's (NQF) draft document titled, "Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement." Boehringer Ingelheim (BI) strongly supports advancements in performance measure development, implementation, and evaluation to improve the delivery, quality, and outcomes of patient care. We believe that performance measures should be consistent with the most current and rigorously-assessed evidence. To this end, we support the NQF's leadership in establishing evidentiary standards for performance measures. We have a few comments on key elements of the report in the following areas:

- Flexibility in Evidence Grading Systems
- Importance of Periodic Evidence Reviews and Measure Maintenance
- Roles of Guideline and Measure Developers
- Transparency in Measure Evaluation Criteria
- Importance of Outcomes Measures and Methodology Transparency
- Guiding Principles Clarification

Maintaining Flexibility in Use of Evidence Grading Systems and Ensuring Transparency

BI applauds NQF's recommendations to ensure that all performance measures are grounded in solid clinical evidence and that the depth and breadth of that evidence does not vary significantly across measures. The report identifies the most relevant and highly-regarded evidence grading systems, GRADE[1] and USPSTF[2], as "preferred" systems for performance measure developers to use in the near term. Many professional societies that utilize grading systems currently use GRADE, USPSTF, or related abridged versions. As such, NQF's recommendation that measure developers use these grading systems is consistent with activities currently underway in the guideline and measure development enterprises. BI supports this approach, as it provides guidance to measure developers but does not limit their ability to use other assessment methods or evidence grading systems.

[1] Grades of Recommendation Assessment, Development and Evaluation (GRADE) introduced in 2004.
Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, et al. (2004) Grading quality of evidence and strength of recommendations. BMJ 328: 1490.

[2] U.S. Preventive Services Task Force Procedure Manual. AHRQ Publication No. 08-05118-EF. Rockville, Maryland: Agency for Healthcare Research and Quality, July 2008. Accessed at www.ahrq.gov/clinic/uspstf08/methods/procmanual.htm

Boehringer Ingelheim

While it is important to ensure that measures are grounded in adequate evidence, it may be difficult for all developers to change to one or two specific grading systems, which is a long-term requirement recommended in the report. Approaches to reviewing evidence are not consistent across guideline developers, so it may be especially challenging for measure developers who use clinical guidelines with different grading systems than those required by NQF to meet the new requirements. This proposed requirement could result in significant measure gaps in certain disease areas (depending on the grading approach of the given professional society focused on that disease condition). If NQF proceeds with establishing *required* grading systems for measure developers, BI requests that the time period to transition from the "preferred" to the "required" grading systems be clearly articulated to measure developers. This transparency will allow developers sufficient time to provide meaningful feedback to NQF on the different grading schemes and fully prepare for the NQF-required approach.

Periodic Evidence Review and Measure Maintenance are Critical

BI supports NQF's proposed requirement for the periodic review of evidence that supports a performance measure when the measure is up for endorsement maintenance. Endorsement and measure maintenance, supplemented with the periodic review of the evidence base, are vital to ensure that measures are consistent with the most recent clinical evidence. Moreover, it is important that measures account for new treatments, technologies, and innovation, ensuring patient access to the best treatment available to improve outcomes.

Guideline Developers, NOT Measure Developers, are Well-positioned to Conduct Evidence Reviews

While BI supports the concept that all measures should be evidence-based, we do not agree that measure developers, in addition to guideline developers should be responsible for the in-depth assessment and grading of the evidence. It is understood that guideline developers, by and large, have the expertise of clinical medical advisors and the resources of skilled staff in reviewing and grading evidence. Conversely, it is unclear whether measure developers have the appropriate experts or adequate resources to hire experts or fund comprehensive evidence reviews.

There are several advantages to maintaining the current approach to measure development. First, basing measures on the strongest recommendations in clinical guidelines helps ensure broad stakeholder consensus around those measures. Further, evidence grading and reviews by both guideline and measure developers, as proposed in this report, will result in significant duplication of efforts for resource-strapped organizations. As such, BI suggests that performance measure development remain a sequential process. Guidelines should be generated or updated first. Performance measures based on key elements of those guidelines should come later. Measure developers should work collaboratively with guideline developers to ensure that the resulting measures exhibit the most desirable attributes, e.g. feasible, valid, and based on recent, reliable clinical evidence.

In addition, given the significance of the proposed changes to the endorsement requirements, BI suggests that NQF consult measure and guideline developers on the additional expectations. From our perspective, published statements by measure developers and guideline developers indicate that these groups are not entirely aligned with the new requirements and responsibilities that NQF proposed in the draft report. For example, the American Medical Association-convened Physician Consortium for Performance Improvement (AMA PCPI) released a position statement regarding the necessary evidence base for measures in June 2009. While AMA PCPI and NQF agree on the need for consistent assessment of evidence, the AMA PCPI's position statement indicates that they believe it is role of the guideline developers, not the measure developers, to conduct and present more comprehensive assessments of the evidence. The AMA PCPI has developed a system to support measure developers in evaluating the rigor and strength of the clinical guidelines prior to developing measures based on them. BI recommends that NQF work closely with measure and guideline developers to clearly understand and appreciate their roles.

Boehringer Ingelheim

Increased Transparency around Measure Evaluation Criteria

The NQF measure evaluation criterion, importance to measure and report, has three sub-criteria[3] whose relative weights were not articulated until this draft report. BI appreciates NQF's proposal to provide additional guidance around the relative importance of the sub-criteria. Specifically, the Task Force recommends that all three sub-criteria be met prior to further evaluation. BI agrees with the proposed all-or-none requirement for the three sub-criteria. A three-pronged approach will make a solid and valid case for why a measure should be developed, endorsed, and implemented. If implemented individually, the criterion may unfairly position one disease condition, process of care, or outcome above another without fully understanding its merit.

Recognizing the Importance of Outcomes Measures and Increasing Transparency around Methodologies

BI agrees with NQF's hierarchy of measure types: outcomes, process, and then structure. Improving patient outcomes should be the main focus and goal of performance measurement and improvement efforts. The ability to quantify patient outcomes through measurement is critical to understanding the impact of various interventions. While the importance of outcomes measures is well-known, not all approaches to measuring and evaluating outcomes have substantial evidence to support their use. Given this challenge, BI recognizes the need for NQF to maintain a certain degree of flexibility while considering outcomes measures (relative to other types of measures).

However, BI recommends that NQF require complete transparency around the methodologies used to develop outcomes measures, as it is essential in determining the appropriateness of the conclusions drawn. Moreover, such data often forms the foundation of future improvement interventions, thus making it imperative that all elements of the methodology are clear and publicly available.

Grounding Measures for Transparency and Value-based Purchasing Initiatives in Strong Evidence

In "The Changing Environment" section of the report, NQF recognizes the importance of grounding measures used in both public reporting and value-based purchasing programs in solid evidence. However, later in the report, NQF only focuses on that which is used for public reporting. With the increased focus on payment reform in the public and private sector and the establishment of the Center for Medicare & Medicaid Innovation by the Affordable Care Act, BI encourages NQF to further acknowledge as part of the guiding principles for this report the importance of ensuring that measures used in value-based purchasing initiatives are grounded in evidence.

Conclusion and Next Steps

In conclusion, we look forward to working together to strengthen the evidence base for measures, which is an important step in improving patient outcomes.

Sincerely,

Hemal Shah, PharmD
Executive Director, Health Economics and Outcomes Research
Boehringer Ingelheim Pharmaceuticals, Inc.
900 Ridgebury Road / PO Box 368
Ridgefield, CT 06877-0368

[3] The three sub-criteria include: high impact; opportunity for improvement; and evidence for the focus of measurement.

AHIP Comments on Measure Evaluation Criteria

AHIP supports the NQF efforts to provide measure review panels with additional guidance on how to evaluate the strength of evidence when considering the importance of measures. Adding clarity around the strength of evidence supporting measures will enable greater continuity in how measure review panels conduct their work. AHIP supports the NQF's subsequent revisions to their measure evaluation criteria based on the report's recommendations. We offer the following specific comments on the report.

Quantity of Evidence

The report lists a general guideline of how many studies would equal a sufficient body of evidence to support a measure. The recommended number of studies per evidence category appears reasonable (5+ studies for High, 2-4 studies for Moderate, etc). AHIP would like additional assurance that measure developers could not game the system by supplying references to a limited set of studies. For example, a measure submission may list five randomized controlled trials support their measure, but there may be newer studies with contradictory evidence. It will be important for NQF to consider the most current studies as well as those with the greatest rigor. NQF should clarify how the veracity of the measure submission will be assessed.

Quality of Evidence

We support the NQF's quality of evidence categories, which mirror those used by GRADE, the U.S. Preventive Services Task Force, and AHRQ's Evidence-based Practice Centers.

Revisions to the Measure Evaluation Criteria

AHIP supports revisions to the measure evaluation criteria that require a measure to meet the all of the following – the measure represents a high impact area or national priority, demonstrates an opportunity for improvement, and is supported by an adequate evidence base. Previously, measures only had to meet one of these criteria. We believe this revision will improve the value of endorsed measures.

Unintended Consequences

The NQF should consider the impact of changing the measure evaluation criteria on measures that have already been endorsed. NQF should also consider the impact on measure developers. More stringent criteria may impact the capacities and breadth of measure developers that have the resources to submit measures to NQF.

CHILDBIRTH CONNECTION
since 1918
Helping women and health professionals make informed maternity care decisions.

June 28, 2010

Helen Burstin, MD, and Karen Pace, PhD
National Quality Forum
601 13th Street NW, Suite 500 North
Washington DC 20005

Dear Helen and Karen:

Thank you for the opportunity to review the draft *Guidance for Evaluating the Evidence Related to the Focus of Quality Measurement*. We appreciate the time and effort put forth by NQF staff and task force members to develop this important guidance and offer the following comments for your consideration.

We are concerned that some statements in the document are out of step with established standards of evidence as we understand them from Childbirth Connection's decade-long work translating results of the wealth of systematic reviews in our field and commissioning or carrying out others to fill gaps.

Expert opinion (report pp. 9 15, 16, 17, 22, 30) is widely understood to be unreliable and is classified at the lowest level of evidence. The GRADE Working Group makes this clear (Guyatt et al. *BMJ* 336:924-26), and the USPSTF Grade Definitions web page does not mention "expert," "opinion" or "consensus." We agree that a high bar should be set for NQF endorsement. Expert opinion does not meet the level described in the new report. Measures based on uncertainty associated with expert opinion are unfair to the stakeholders implementing them. Future disproving evidence would embarrass NQF. For clarity, NQF's guidance for developers, Steering Committees and TAPs should echo the GRADE Working Group on this matter

Guidelines. Methods and standards for guidelines development vary widely in the U.S., and many fail to measure up to evolving international standards. Until we have and implement results of the IOM guidelines standards report (and assuming that the report's recommendations meet the bar that NQF is now setting), it is inappropriate to assume that a guideline recommendation from a well-known national organization is based on a formal evaluation of the weight of the best available evidence, as the report expects and as is routinely done by, e.g., NICE in the U.K. We feel that NQF should not depend for now on guideline developers for rigorous search, identification and assessment of a body of evidence. The cited Lohr presentation, Grilli et al. study, and Tricoci et al. study and similar findings in other fields (e.g., Chauhan et al. *Am J Obstet Gynecol* 2006(194):1564-72) call for caution. The meaning of current ratings is generally unclear. We do not believe that the caveat in lines 277-79 covers these limitations. Until we know what the proposed national standards are, we also cannot be sure that meeting them (lines 275-76) will be adequate for the standards to which NQF now aspires. The second column in Table 7 carries over many items relating to guidelines that are now inconsistent with NQF's bar.

281 Park Avenue South, 5th Floor
New York, NY 10010
p 212.777.5000
f 212.777.9320
www.childbirthconnection.org
info@childbirthconnection.org

All studies vs. selected studies (report pp.15, 22, 23). Due to the proliferation of studies and information overload, it would not be possible to review all studies on a specific topic. Nor is it desirable, as most studies are not trustworthy guides for practice. The international evidence standard is to include all studies that meet a priori criteria for relevance and quality. The latter is generally about type of design and whether the study scores well enough on a checklist to eliminate weaker studies.

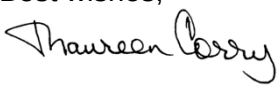
Developers as reviewers, and systematic reviews. To carry out protections against bias, systematic reviewing requires skill, experience, time and resources that are beyond the scope of many measure developers. Conventional narrative reviewing has few rules and is much easier. However, it is highly vulnerable to bias and thus considered to be at the lowest level of evidence, and not useful as a guide for practice. Developers should be encouraged to base their measures whenever possible on up-to-date well-conducted systematic review(s) that others have vetted and published. This is stated on line 273, but is not clear in Tables 4 and 7 and in the statement in lines 132-33 and 137-39 that measure developers or designates should carry out evidence reviews. Many thousands of systematic reviews are available in the journal literature, in databases, and on websites of agencies and organizations throughout the world. The hard work of rigorously clarifying the weight of the better quality evidence to answer thousands of focused clinical and health system questions has in many cases already been done by those with the expertise, interest, and resources. These are underutilized, and should take their place in the continuum of clinical effectiveness products as optimal sources for informing guidelines, performance measures and decision aids. Also worthy of support are Overviews of Systematic Reviews (covered in *Cochrane Handbook for Systematic Reviews of Interventions*) and Best Evidence Reviews (e.g., *Clinical Evidence*).

Patient experience of care. We agree that the evidence in support of patient experience measures differs, and have the impression that data for CAHPS survey development and testing have been extensive and impressive. Guidance on evidence for patient experience of care measures is needed in Table 7.

Harms are generally inadequately included and reported in primary data research and are thus difficult to address well in the systematic reviews, guidelines, quality measures, and decisions aids that are derived from the original studies. NQF-endorsed quality measures have the potential to increase the use of the structures and processes that are the topics of the measures. To avoid unintended consequences, it is appropriate to ask developers, Steering Groups and TAPs to look carefully at the “included studies” and also consider other data sources that might shed light on the safety of practices that they consider recommending for endorsement. The thoughtful chapter on Adverse Effects in the *Cochrane Handbook for Systematic Reviews of Interventions* would help point the way toward appropriately addressing harms in the new report.

Please feel free to contact me if I can provide further support for this important document for NQF and our national health care system.

Best wishes,



Maureen P. Corry
Executive Director