Committee report: Method for evaluating conditions nominated for population-based screening of newborns and children

Ned Calonge, MD, MPH[1], Nancy S. Green, MD[2], Piero Rinaldo, MD, PhD[3], Michele Lloyd-Puryear, MD, PhD[4], Denise Dougherty, PhD[5], Coleen Boyle, PhD[6], Michael Watson, PhD[7], Tracy Trotter, MD[8], Sharon F. Terry[9], and R. Rodney Howell, MD[10]; for the Advisory Committee on Heritable Disorders in Newborns and Children

Abstract: The Secretary's Advisory Committee on Heritable Disorders in Newborns and Children is charged with evaluating conditions nominated for addition to the uniform screening panel and consequently making recommendations to the secretary of the US Department of Health and Human Services. This report describes the framework by which the committee approaches its task. Key decision nodes include initial review of every nomination to determine whether conditions are amenable for systematic evidence review, review of systematic evidence reviews conducted by the committee's external review group, and deliberation and formal recommendation for addition or exclusion to the uniform panel. Data analyzed include the accuracy and specificity of screening and diagnostic tests for nominated disorders, the extent of predicted health benefits, harms impact on disease course, and cost from early diagnosis and treatment. The committee process is guided by approaches used by similar entities, but more flexible criteria are sometimes needed to accommodate data limitations stemming from the rarity of many of these conditions. Possible outcomes of committee review range from recommendation to add a nominated condition to the uniform panel; provide feedback on specific gaps in evidence that must be addressed before making a decision; or rejection of a nomination (e.g., because of identified harms). The committee's structured evidence-based assessment of nominated conditions supports a consistently rigorous, iterative and transparent approach to its making recommendations regarding broad population-based screening programs for rare conditions in infants and children. ***Genet Med* 2010:x(x):000–000.**

Key Words: *newborn screening, evidence review, nomination, rare condition, advisory committee*

From the [1]Colorado Department of Public Health and Environment, Denver, Colorado; [2]Department of Pediatrics, Columbia University Medical Center, NY, New York; [3]Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, Minnesota; [4]Department of Health and Human Services (HHS), Health Resources and Services Administration, Rockville, Maryland; [5]HHS, Agency for Healthcare Research and Quality, Rockville, Maryland; [6]HHS, Centers for Disease Control and Prevention, Rockville, Maryland; [7]American College of Medical Genetics, Bethesda, Maryland; [8]San Ramon Valley Primary Care Medical Group, San Ramon, California; [9]Genetic Alliance, Washington, DC; and [10]Department of Pediatrics, Miller School of Medicine, University of Miami, Miami, Florida.

Nancy S. Green, MD, Columbia University Medical Center, 630 West 168 St., Black Building 2-241, New York, NY 10032. E-mail: nsg11@columbia.edu.

Disclosure: The authors declare no conflict of interest.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Agency for Healthcare Research and Quality, the Centers for Disease Control and Prevention, the Health Resources and Services Administration, or the U.S. Department of Health and Human Services.

Submitted for publication September 24, 2009.

Accepted for publication January 5, 2010.

DOI: 10.1097/GIM.0b013e3181d2af04

Because of the historical heterogeneity of state-based legislation and processes, conditions targeted by newborn screening programs had varied widely between states.[1,2] The Secretary's Advisory Committee on Heritable Disorders in Newborns and Children (the Advisory Committee) was chartered in 2003 under Section 1111 of the Public Health Service Act, 42 U.S.C. 300b-10, as amended in the Newborn Screening Saves Lives Act of 2008. The Advisory Committee has, among its charges, the responsibility of making evidence-based recommendations at the national level regarding important health conditions for which newborns and children should be screened. Based on recommendations from the American College of Medical Genetics,[3] the Advisory Committee generated a recommended uniform screening panel. Subsequently, the Advisory Committee has begun to evaluate and update that panel.[4] Committee recommendations are based on a standardized process it has developed for evaluating disorders nominated for inclusion in the universal screening panel, weighing the evidence and gathering input from key stakeholders.[4] Key aspects across the processes for nomination and evaluation for the proposed conditions are transparency, broad accessibility, rigorous evidence-based approaches, and consistency. This article outlines the Advisory Committee's process for evaluating the reports submitted by the Advisory Committee's External Review Workgroup (ERW); and for making subsequent recommendation(s) to the secretary, US Department of Health and Human Services (HHS) regarding changes to the uniform panel. To date, the committee process has been focused on newborn screening, but the process may be applied to other population-based screening of children for rare disorders.

The Advisory Committee process builds on a broad array of methodologies for systematic evidence reviews (SERs) including approaches put forth by the World Health Organization,[5] the National Academies of Science,[6] the Council of Regional Genetic Networks,[1] the American Academy of Pediatrics Newborn Screening Task Force,[2] the Centers for Disease Control and Prevention's Analytic validity; Clinical validity; Clinical utility; and Ethical, legal, and social implications (ACCE),[7] the United States Preventive Services Task Force,[8] the Evaluation of Genomic Applications in Practice and Prevention,[9,10] and the workgroup the American College of Medical Genetics Newborn Screening Expert Panel.[3] None of these entities is empowered to enforce adoption of its recommendations by any entity, but the recommendations have had substantial impact on clinical practice and public health.[11] Unlike other deliberative groups, the work of this committee differs from other approaches in its nearly exclusive focus on screening newborns and children for rare conditions. In addition, only the Advisory Committee and the United States Preventive Services Task Force are authorized by federal legislation to make recommendations to the secretary of HHS.

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

The Advisory Committee uses the following process to complete the evaluation of an important health condition nominated to make its recommendations (Ref. 4 and described in the minutes posted the Advisory Committee website www.hrsa.gov/heritabledisorderscommittee):

1. A condition is nominated for consideration via a structured nomination process. The executive secretary of the Advisory Committee oversees an administrative review for completeness of the nomination via the process outlined at www.hrsa.gov/heritabledisorderscommittee/nominate.htm.
2. Once complete, nomination package is assessed by the Committee's internal Nomination and Prioritization Workgroup for the likelihood of sufficient information to conduct a SER on the natural history and severity of the condition, the analytical and clinical validity of the screening test(s), and effectiveness of treatment. Information and citations in the nomination package comprise the starting point for the Workgroup's assessment of these issues. At this point, cost is not a consideration. Following a report from the Workgroup and discussion, the full committee votes on whether a nominated condition should move forward for a full evidence review.
3. If the Advisory Committee agrees to move the nomination forward, the nomination package is assigned to the Committee's ERW for a SER. The ERW was established through a competitive contract process through Health Resources and Services Administration to solicit establishment of an evidence review team for the Advisory Committee and works independently of the committee to make its own objective assessment of the evidence base for key topics related to the nominated condition (described in Ref. 4).
4. Once a SER is completed, an EWG draft report submitted to the Decision Process Workgroup and to the full Committee. Both an author of the SER and the Decision Process Workgroup present an overview of the report at the next committee meeting.
5. After full discussion of the reports, using the process outlined in this document, the committee makes one of the following recommendations to the secretary of HHS:

- Inclusion of the condition in the uniform screening panel for newborns with heritable conditions;
- formulation of specific questions to be addressed by either the EWG or researchers in the field before a committee recommendation can be made;
- recognition of the need for broader evidence before a nomination can be reconsidered;
- recommendation that a condition not be included in the uniform screening panel at this time.

The Advisory Committee Chair sends a letter with the recommendation(s) to the secretary, HHS. Within 180 days of the Advisory Committee issuing the recommendation, the secretary is required to adopt or reject any recommendations and to publicize that determination. Recommendations of the Advisory Committee and the subsequent determinations of the HHS secretary are considered but are not binding on state screening programs. Ultimately, states determine their own screening panels.[1]

This article lays out the analytic framework, related key questions, and criteria for assessing evidence that the committee uses to evaluate SERs. A section on "weighing the evidence" addresses study design and criteria for evaluating study quality and adequacy of evidence. A final section addresses translating evidence into committee recommendations to the secretary.

Key considerations regarding the evidence base for evaluating a nominated condition

The Advisory Committee addresses a set of key questions informed by the analytic framework it has adapted from other sources. Although this approach is similar to those used in other evidence-based recommendation processes, the Advisory Committee recognizes that some allowances likely will need to be made for evaluations involving rare disorders. The rapid pace of development in genetics and screening technologies makes it increasingly feasible to identify children with specific rare conditions much earlier in life than those with the associated clinical presentation. Earlier identification can facilitate timely, effective treatment, and therefore avoid preventable child morbidity and mortality, as well as "diagnostic odysseys." Highly prevalent and relatively well-characterized disorders have been the focus of evidence-based analyses. In contrast, for the rare heritable disorders under consideration by the Advisory Committee, building the evidence base is much more challenging from a scientific and practical standpoint. There will be limited evidence-based assessments on the clinical significance of screening and diagnostic test results, the phenotypic expression of detected genotypes, the full range of potentially effective medical or other management options, and the harms or other benefits that might be associated with testing and subsequent interventions may not be fully understood.

The Advisory Committee recognizes the likelihood of limited available peer-reviewed, large-scale controlled trials using rigorous intervention research designs for evaluation of the rare conditions typically nominated for inclusion in the recommended uniform screening panel. For many if not most disorders, it may be necessary to consider evidence from studies using less robust research designs, such as modest-sized open-label clinical studies for evaluating treatment and population-based observational studies, as available, when evaluating conditions or testing technologies. For example, in February 2009, the Advisory Committee considered an SER of severe combined immune deficiency in which outcomes of strategies for stem cell transplantation were compared with historic rather than randomized controls (http://www.hrsa.gov/heritabledisorderscommittee/meetings/17th meeting). Some of the "gray" literature will also be considered, such as data that have not yet undergone peer review. Despite these limitations, the Advisory Committee has developed a process that allows for objective, replicable decisions based on a sequential decision process. Available approaches to SERs are limited and compounded by the potential for unintended and/or underestimated consequences and costs of implementing new technologies. Despite these limitations, the Advisory Committee has made tangible progress to date and will continue to develop rigorous review approaches relevant to evaluating evidence and making recommendations for screening of newborns and children with heritable disorders.

More information on (A) defining analytic validity, (B) ranking the quality of data sources, and (C) assessing study quality are available on the Committee website, www.hrsa.gov/heritabledisorderscommittee. Review criteria are based on those established by others.[3,7–10]

Evaluation of the external review workgroup report

The Advisory Committee evaluates the SER in three broad areas: the nominated condition (incidence, prevalence, significance); the screening test and diagnostic tests for the condition based on current best available technical approach(s) (clinical utility, analytical and clinical validity); and the treatment(s) (clinical utility, efficacy or effectiveness).[4] Key questions for

© 2010 Lippincott Williams & Wilkins

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

these areas are shown in Figure 1. Based on the SER and additional more recent information that may be available, the Advisory Committee, in open meetings, evaluates whether the current evidence for each of the six key questions is adequate or inadequate. On the basis of the strength of the evidence and the predicted magnitude of net benefit (benefits minus harms), the Advisory Committee makes a specific recommendation regarding the outcome of the nomination.

Key Question 1

Is there direct evidence that screening for the condition at birth leads to improved outcomes for the infant or child to be screened? Are there potential benefits for the child's family?

This is the overarching question for the evidence review. Outcomes encompass the impact(s) of screening, diagnosis or lack of diagnosis, prognosis, therapeutic choice or lack of therapy, patient outcomes, and familial and societal issues. Positive patient outcomes are typically measured as reductions in morbidity or mortality. While of debatable relevance to consideration by the Advisory Committee, the value of screening may extend to aspects such as improved quality of life and patient or family satisfaction with health and related services for the condition.

If adequate direct evidence—by one or more high-quality studies of randomized trial or definitive population-based study design—is available to make a recommendation, there is no need to address the remaining key questions in the analytic framework.

Key Question 2

Is there a case definition that can be uniformly and reliably applied? What is the incidence and prevalence of the condition? What are the natural history and the spectrum of disease of the condition, including the impact of early recognition and treatment versus later recognition and delayed or no treatment?

For each nominated condition, an agreed on case definition based on rigorous criteria is essential for ascertaining incidence, prevalence, and severity. Screening for a condition of lower clinical severity can be deemed important due to a high incidence, likewise a rare condition can be important due to serious health consequences. Understanding the spectrum of disease is essential in considering whether there are cases of the condition for which treatment is not effective or otherwise unwarranted or if the condition is readily clinically identified in a newborn or child without screening.

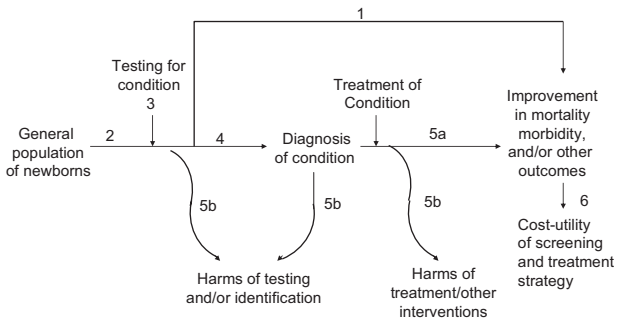

Fig. 1. The analytic framework depicts the considerations of evidence for population-based screening of newborns for a specific important health condition (or set of conditions). Each number corresponds to a key question which, in total, describes the structured analysis for considering the existing data (Adapted from U.S. Preventive Services Task Force Procedure Manual, http://www.ahrq.gov/clinic/uspstf08/methods/procmanual.pdf.)

Key Question 3

Is there a screening test or screening test algorithm for the condition with sufficient analytic validity?

Analytic validity refers to the technical accuracy of the laboratory test in measuring the intended analyte(s), as distinguished from clinical validity, which is the ability of the test or test algorithm to predict the development of clinical disease. Evaluation of the evidence for sufficient analytic validity is an assessment of the sensitivity and specificity of the testing protocol for detecting a target disorder or set of disorders. Analytic validity includes preanalytical, analytical, and postanalytical issues, as well as the feasibility of standardization between different laboratories performing the same test—i.e., the reliability. The four specific elements of analytic validity of the test(s) include analytic sensitivity (or the rate of analytic detection), analytic specificity, laboratory quality control, and assay robustness. Analytic sensitivity defines how effectively the test identifies specific analytes present in a clinical sample. Analytic specificity defines how effectively the test correctly classifies samples that do not have specific analytes. Quality control assesses the procedures for ensuring that results fall within specified limits or cutoff values. Robustness measures how resistant the assay is to changes in preanalytic and analytic variables.

The Advisory Committee's goal is that testing programs across the country be able to implement testing or testing platforms with comparable levels of analytic validity. For example, using tandem mass spectrometry applied to the uniform panel, there is growing evidence suggesting that this laboratory methodology could achieve a detection rate between 1:2000 and 1:3000, a false-positive rate $<0.3\%$, and a positive predictive value $>20\%$ for conditions within the uniform panel.[12]

Key Question 4

Has the clinical validity of the screening test or screening algorithm, in combination with the diagnostic test or test algorithm, been determined and is that validity adequate?

The clinical validity of a test (or test algorithm) defines its ability to detect or predict the associated disorder (phenotype). There are two parts to the question of clinical validity:

1. Is the evidence sufficient to conclude that the clinical validity is known? This involves only a consideration of the adequacy (strength and quality) of the evidence in the SER that we know the sensitivity and specificity of the screening and diagnostic testing or testing algorithm, i.e., its ability to accurately detect the disorder.
2. Is the identified level of clinical validity sufficient to justify testing?

This question gauges the ability of the screening test (or test algorithm), when used to identify individuals who merit diagnostic testing, to detect as many as possible affected individuals who manifest clinical disease, and to minimize the occurrence of false positives. These issues relate to both performance of the screening and diagnostic tests and the incidence of the condition. Consideration must be given to the potential for individuals to test positive but not develop clinical disease (those who screen positive and whose disease is confirmed by diagnostic testing but who do not develop signs or symptoms). Rare conditions may exhibit a wide range of signs and symptoms, so

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

that defining phenotype may include a range of disease manifestations. No test is perfect, requiring tradeoffs among false positives, false negatives, and identification of nonclinical conditions also all impact clinical utility (see key question 5).

It is possible that evidence on clinical validity will be adequate, whereas evidence on analytic validity is not adequate. Under unusual circumstances, such as where the likelihood both of clinical benefit is exceptionally high and the harms are low, the Advisory Committee might make a positive recommendation to add the condition to the uniform panel, although issues of dissemination and implementation will need to be carefully considered.

Key Question 5

What is the clinical utility of the screening test or screening algorithm?

a. What are the benefits associated with use of the screening and diagnostic tests and the treatment?
b. What are the harms associated with screening, diagnosis, and treatment?

The clinical utility of a test defines the elements of both testing and treatment that need to be considered when evaluating the benefits and harms or risks associated with its introduction into routine practice. In considering benefits, the question of clinical utility involves the ability of screening to lead to improved important health outcomes, primarily decreased morbidity and mortality. Broader benefits to the individual infant, such as nonclinical interventions or benefits to family and community, such as avoiding a diagnostic odyssey or informing nonmedical decision making, may also be considered.

The consideration of harms or risks includes evaluating the potential for risks of physical harm associated with testing, identification, and/or treatment as well as those harms or risks that are nonphysical, such as the possibility for stigmatization, unnecessary anxiety, adverse impacts on parent and family relationships, and other ethical, legal, and social implications. Risk of physical harm is an aspect inherent to virtually all medical intervention; evaluation requires an assignment of an estimate of the potential morbidity or even mortality to support decisions regarding net benefit of testing and treatment when compared with treatment outcome subsequent to clinical diagnosis.

Questions to evaluate clinical utility for testing include does the screening test result, once validated by diagnostic testing, inform valid clinical decision making? Can the diagnosis be made in an accurate and timely manner? How likely will screening lead to the prevention or amelioration of adverse health outcomes associated with the disorder (assuming the adoption of an accompanying efficacious treatment conditioned on test results)? Have the risks and benefits associated with the introduction of testing for this condition been identified (again, assuming the adoption of an accompanying efficacious treatment conditioned on test results)? Are quality assurance assessment procedures in place for controlling preanalytic, analytic, and postanalytic factors that could influence the risks and benefits of testing? Have pilot trials assessed the performance of testing under real-world conditions? Are there practical limits to the use or availability of the screening or diagnostic tests, such as patent or licensing protections or limiting capacity for diagnostics?

When considering treatment, the question of clinical utility involves evaluating whether treatment of the condition detected through screening improves important health outcomes when compared with delays until clinical detection, based on available treatment. Health outcomes may encompass the impact(s) of diagnosis or lack of diagnosis, the prognosis, therapeutic choice or lack of therapy, the patient outcome, and familial and societal issues. These outcomes are not of equal weight or value and involve balancing the tradeoffs between different favorable and unfavorable outcomes. Other questions regarding treatment include are treatment protocols for affected children standardized, widely available, and if appropriate, FDA approved? Are there subsets of affected children more likely to benefit from treatment that can be identified through testing or clinical findings? It is important to note that treatment may include a broad list of interventions including counseling and support services, beyond the conventional definition of medical therapy.

The consideration of both potential risk and benefit is crucial to enabling the Advisory Committee to balance factors influencing the recommendation for screening for a condition.

Key Question 6

How cost effective is the screening, diagnosis, and treatment for this disorder compared with usual clinical case detection and treatment?

Cost effectiveness refers to the ratio of a procedure's cost to its effectiveness. Cost effectiveness can be assessed using a variety of templates: medical costs and benefits alone, direct plus immediate indirect costs, and societal-level costs. Benchmarks for whether a test or treatment is cost-effective shift and depend on the perspective taken as well as the method used in the analysis, among other things.

Peer-reviewed published evidence on the cost effectiveness of most health care services is limited, and thus studies involving primary data collection on comprehensive costs or cost effectiveness related to newborn or child screening and treatment for rare conditions would not be expected. Consideration of cost effectiveness for screening should include available data on the incremental costs for screening, diagnosis, and treatment for a disorder, compared with costs for not screening. The approaches used by Carroll and Downs[13] will serve to guide the Advisory Committee's analysis of the impact of cost of screening, diagnosis, and treatment for a particular condition.

Weighing the evidence

Study Design

SER methodology requires that consideration of the strengths and weaknesses of study design be used to provide a quality ranking of data sources on treatment of identified conditions (see Appendices), while considering the potential limitations in the quality of some of the data sources, as described above. Criteria for assessing the quality of study design differ for different kinds of study questions.[10] For example, for questions of analytic validity of screening tests, the best information comes from collaborative studies using a single large, carefully selected panel of well-characterized control samples that are blindly tested and reported, with the results independently analyzed.[1] Data from proficiency testing schemes can provide information about all three phases of analytic validity (i.e., analytic, pre-, and postanalytic) and interlaboratory variability.

Criteria for Evaluating Study Quality

The assessment of the quality of data from studies to potentially be included in an SER includes evaluating the number of reports, the quality of study design, the total number of studies, subjects and treatment arms, the numbers of positive and negative controls studied, and the range of study methodologies

 © 2010 Lippincott Williams & Wilkins

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

represented. The consistency of findings will be assessed formally (e.g., by testing for homogeneity when possible) or by less formal methods (e.g., providing a central estimate and range of values) when sufficient data are lacking. One or more internally valid studies do not necessarily provide sufficient information to justify large-scale public health implementation. Support for screening for a condition or use of a test in universal public health applications generally requires studies that provide estimates of analytic validity and effectiveness of early treatment that are appropriate to use in diverse "real-world" settings. Also, existing data may support the reliable performance of one methodology, but no data may be available to assess the performance of one or more other methodologies.

Evaluating the Adequacy of Evidence

The adequacy of the evidence to answer each of these key questions can be summarized and then classified across the questions as adequate or inadequate.[12,14,15] This is also referred to as assessing the strength of the linkages in the chain of evidence. To support a recommendation, adequate evidence would require studies of fair or better quality for key question 5 and evidence to meet satisfactorily most if not all of the other key questions. A determination of "insufficient evidence" could be based on the absence of evidence, studies of poor quality, or studies with conflicting results.

The evidence is examined overall and a decision is made regarding whether the evidence is graded as overall adequate or inadequate to answer the key question.

- When the quality of evidence is adequate, the observed estimate or effect is likely to be real, rather than explained by flawed study methodology, and the Advisory Committee concludes that the results are unlikely to be affected strongly by the results of future studies, all other things being equal.
- When the quality of evidence is inadequate, the observed results are more likely to be the result of limitations and/or flaws in study methodology rather than an accurate assessment, and subsequent information is more likely to change the estimate or effect enough to change the conclusion.
- Availability of only marginal quality studies always results in inadequate quality.

Magnitude and certainty of net benefit

Essential factors for the development of a recommendation include

- The relative importance of the outcomes considered;
- the health benefits associated with testing for the condition and subsequent interventions; or, if the actual or estimated health benefits are not available from the literature, then the maximum potential benefits;
- the harms associated with testing for the condition such as adverse clinical outcomes, increase in risk, unintended ethical, legal, and/or social issues that result from testing and subsequent interventions; or, if the actual or estimated harms are not available from the literature, then the maximum potential harms; and
- the efficacy and effectiveness of testing for the condition and follow-up compared with current practice, which might even include no specific medical intervention. Benefits and harms may include psychosocial, familial, and social outcomes.

Consistent with the processes of other evidence-based recommendation groups, the magnitude of net benefit (benefit minus harm) can be graded as at least moderate, small, or absent (zero or net harm). For the purposes of the Advisory Committee in making recommendations, moderate or greater net benefit will be considered "significant" and will support a recommendation to add the condition, and zero/harmful net benefit will support a recommendation not to add the condition. Those conditions where the magnitude of net benefit is classified as small will be discussed on a case-by-case basis and classified as either significant or not significant. A recommendation to add a condition where testing is expected to provide only small net benefit should be supported by a high degree of certainty based on the evidence (see certainty of net benefit below).

Based on the summaries of the evidence for each key question and the chain of evidence, the certainty of the conclusions regarding the net benefit can be classified as sufficient or low. A conclusion to either recommend adding or not adding the condition with sufficient certainty has an acceptable risk or level of comfort of "being wrong" and thus a low susceptibility to being overturned or otherwise altered by additional research. Insufficient certainty should not lead to a recommendation for or against adding the condition but should lead to a recommendation for further research.

Translating evidence into Advisory Committee recommendations

The process is designed to be streamlined, transparent, evidence based, and consistent throughout the review process and across different conditions under consideration. After the evidence-based review is completed, the Advisory Committee will review the report and put forth a formal recommendation based on the quality and strength of the data as summarized in the evidence review. Additional factors may also be weighed, such as ethical, legal, and public heath issues. When relevant, the Advisory Committee will also consult with other federal advisory committees when developing their recommendations.

Recommendations will be based on the level of certainty that testing will result in significant net health benefit, based on the evaluation of the evidence. Table 1 serves to outline the recommendation category.

Category 1: the committee has sufficient certainty of significant net benefit to recommend adding the condition to the uniform panel.

Category 2: the evidence is insufficient to make a recommendation.

- However, there is compelling potential for net benefit, and the committee wants to make a strong recommendation for additional studies to fill in the evidence gaps.

Category 3: the evidence is insufficient to make a recommendation.

- There is insufficient evidence of potential net benefit to lead the committee to want to make a strong recommendation regarding additional studies.

Category 4: the committee has sufficient certainty of no net benefit, or of net harm, to recommend not adding the condition to the uniform panel at this time.

Conditions given a Category 2 recommendation deserve special comment. These are conditions where the evidence is inadequate to reach a conclusion and make a recommendation based on at least fair evidence of clinical utility and significant net benefit but for which, there are contextual issues that support a strong recommendation for pilot studies to fill in the gaps in evidence as quickly as is feasible. Contextual issues might include known benefits associated with testing (and intervention) for similar conditions, a high

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

Table 1 Decision matrix for Advisory Committee recommendations

Category	Recommendation	Level of certainty	Magnitude of net benefit
1	Recommend adding the condition to the uniform panel	Sufficient	Significant
2	Recommend not adding the condition, but instead recommend specific additional studies	Insufficient, but the potential for net benefit is compelling enough to recommend specific additional studies to evaluate	Potentially significant, and supported by contextual considerations
3	Recommend not adding the condition based on current knowledge	Insufficient, and substantial additional evidence is needed to make a conclusion about net benefit	Unknown
4	Recommend not adding the condition to the uniform panel	Sufficient	Zero or net harm

incidence that would translate to potential substantial net benefit, the availability of promising but yet unproven new therapies, or indirect evidence of perhaps lower value health outcomes but with evidence of low potential harm. For these conditions, the Advisory Committee will encourage the undertaking—and funding—of one or more specific studies to address key knowledge gaps and/or evaluate specific aspects of case definition, screening and/or treatment for which some uncertainty persists. For example, one or more population-based pilot studies that are applicable to heterogeneous US populations may need to be performed and evaluated before the Advisory Committee making any decision about inclusion or exclusion in newborn screening. Conditions for which specific data are needed should be reevaluated at a time when sufficient new data exist that may be available to fill in the gaps in the chain of evidence. The decision whether to refer for pilot studies should be made with careful considerations of the potential harms associated with the premature acceptance of unproven clinical strategies, weighed against the potential but health benefits and potential harms of waiting for more compelling evidence.

Despite creating a rigorously standardized method for evaluating the evidence, how the weight of evidence is interpreted by Advisory Committee members—and others—may vary depending on differing values and perspectives. Variation in interpretation is inherent in any group decision-making process and may be more pronounced when evidence is limited to nonrandomized experience with screening or intervention. Transparency throughout the process, including disclosure of differing interests, is intended to minimize subjective influences. The process described herein a relatively new approach for the Advisory Committee and has been used only a few times and may evolve with experience gained from considering nominated conditions.

Advisory Committee recommendations to the HHS secretary are accompanied by:

- summary of evidence and strength of recommendation(s);
- recommendation(s) of other professional groups;
- discussion of rationale for Advisory Committee recommendation(s) that will explicitly state the basis on which the recommendations were made, i.e., a sufficient body of evidence based on results of controlled trials, observational studies, case series, expert opinion, focus groups, cost-effectiveness analysis, policy analyses, ethical analysis, and other inputs; and
- recommended subsequent surveillance, research, education, and program evaluation activities, if applicable.

The Advisory Committee's recommendations are intended to provide transparent, authoritative advice. These may also be used to promote specific research to fill in gaps in the evidence for specific conditions. Three elements, discussed in detail above, are considered in making recommendations:

1. The magnitude of net benefit (are the benefits of screening, diagnosis, and treatment minus the harms significant?)
2. Overall adequacy of evidence (does the evidence overall meet the standards for having adequate quality?), and
3. Certainty of net benefit/harm (is the committee sufficiently certain that the research supports a conclusion that benefits exceed harms or not?).

REFERENCES

1. Therrell BL, Hannon WH. National evaluation of US newborn screening system components. *Ment Retard Dev Disabil Res Rev* 2006;12:236–245.
2. Serving the family from birth to the medical home. Newborn screening: a blueprint for the future—a call for a national agenda on State Newborn Screening Programs. *Pediatrics* 2000;106(2 Pt 2):389–422.
3. Newborn screening: toward a uniform screening panel and system. *Genet Med* 2006;8(suppl 1):S1–S252.
4. Green NS, Rinaldo P, Brower, et al. Committee report: advancing the current recommended panel of conditions for newborn screening. *Genet Med* 2007; 9:792–796.
5. Wilson JM, Junger G. Principles and practice of screening for disease. Geneva: World Health Organization, 1968.
6. National Research Council. Committee for the study of inborn errors of metabolism. Genetic screening: programs, principles and research. Washington, DC: National Academy of Sciences, 1975.
7. ACCE. Available at: http://www.cdc.gov/genomics/gTesting/ACCE.htm. Accessed February 7, 2009.
8. United States Preventive Services Task Force. Available at: http://www.ahrq.gov/clinic/uspstmeth.htm. Accessed February 7, 2009
9. Evaluation of Genomic Applications in Practice and Prevention (EGAPP). Available at: http://www.egappreviews.org/about.htm. Accessed February 7, 2009.
10. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:15–20.
11. Merenstein D, Daumit GL, Powe NR. Use and costs of non-recommended tests during routine preventive health exams. *Am J Prev Med.* 2006;30:521–527.
12. Rinaldo P, Safari S, Tortorelli S, Matern D. Making the Case for Objective Performance Metrics in Newborn Screening by Tandem Mass Spectrometry. *Ment Retard Dev Disabil Res Rev* 2006;12:255–261.
13. Carroll AE, Downs SM. Comprehensive cost-utility analysis of newborn screening strategies. *Pediatrics* 2006;117:S287–S295.
14. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D; U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med* 2007;147:871–875.
15. West S, King V, Carey T, et al. Systems to rate the strength of scientific evidence. *Evid Rep Technol Assess* 2002;47:1–11.

 © 2010 Lippincott Williams & Wilkins

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

APPENDICES

Appendix A: Defining analytic validity

Preanalytic phase: such as sample stability and reagent performance.

Analytic phase: evaluating accuracy (including method comparison), precision (both interassay and intra-assay), recovery, linearity, carryover (if applicable), detection limits, signal suppression (if applicable, especially for tandem mass spectrometry), intensity criteria (signal/noise), age- and gender-matched reference values (if applicable), disease range, and cutoff level defining clinical significance (required for second-tier test).

Postanalytic phase: include evaluation of interpretive guidelines, used to define a case, the spectrum of differential diagnoses, and the algorithm for short-term follow-up/confirmatory testing (biochemical, in vitro, and/or molecular).

Appendix B: Ranking the quality of data sources

Hierarchy of study designs/data sources, based on generally accepted principles of evidence review

Level 1—usually yields good quality evidence.

- Collaborative study using a large panel of well-characterized samples.
- Summary data from well-designed external proficiency testing schemes or interlaboratory comparison programs.
- Well-designed longitudinal cohort studies.
- Meta-analysis of randomized controlled trials.
- Randomized open-label clinical studies.

Level 2—usually yields fair quality evidence.

- Other data from proficiency testing schemes.
- Well-designed peer-reviewed studies (e.g., method comparisons, validation studies) or case-control or cohort studies.
- Expert panel reviewed FDA summaries.
- A single randomized controlled trial.
- Nonrandomized open-label clinical studies.

Level 3—depending on flaws may yield fair or poor quality evidence.

- Less systematic peer-reviewed studies, case-control studies.

Level 4—usually yields poor quality evidence.

- Unpublished and/or nonpeer-reviewed research, clinical laboratory or manufacturer data.
- Studies on performance of the same basic methodology, but used to test for a different target.
- Case series.

Level 5—usually yields poor quality evidence.

- Consensus guidelines.
- Expert opinion.

Appendix C: Assessing study quality

1. Clear description of test or disorder/phenotype and outcomes of interest.
2. Adequate description of study design and methodology.
   - For test evaluation: specific methods evaluated; number of positive samples and negative controls tested.
   - For clinical validity: clear description of clinical outcomes.
   - Was data collection retrospective or prospective?
   - Were subjects randomized?
   - Were intervention and evaluation of outcomes blinded?
3. Interventions are clearly identified, scientifically sound, and consistently provided.
4. Adequate descriptions of the basis for the "right answer."
   - Comparison to a "gold standard" referent test.
   - Consensus (e.g., external proficiency testing).
   - Characterized control materials (e.g., National Institute of Standards and Technology*, sequenced).
5. Avoidance of biases.
   - Blinded testing and interpretation.
   - Specimens represent routinely analyzed clinical specimens in all aspects (e.g., collection, transport, processing).
   - Reporting of test failures and uninterpretable or indeterminate results.
6. Analysis of data.
   - Is the information provided sufficient to rate the quality of the studies?
   - Are the data relevant to each outcome identified?
   - Is the analysis or modeling explicit and understandable?
   - Are analytic methods prespecified, adequately described, and appropriate for the study design?
   - Were losses to follow-up and resulting potential for bias accounted for?
   - Is there assessment of other sources of bias and confounding?
   - Are there point estimates of impact with 95% CI?
   - Is the analysis adequate for the proposed use?

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.