

An evidence development process for newborn screening

James M. Perrin, MD^{1,2}, Alixandra A. Knapp, MS¹, Marsha F. Browning, MD, MPH^{2,3},
Anne Marie Comeau, PhD⁴, Nancy S. Green, MD⁵, Ellen A. Lipstein, MD, MPH¹,
Danielle R. Metterville, MS¹, Lisa Prosser, PhD⁶, Denise Queally, JD⁷, and Alex R. Kemper, MD, MPH⁸

BACKGROUND

Challenges in the evidence review for rare diseases

The US Health and Human Services Secretary's Advisory Committee on Heritable Disorders in Newborns and Children (AC), initially convened in 2004, began work at that time to determine which conditions to recommend for a national minimum uniform newborn screening panel. The role of the AC is to make recommendations to the Health and Human Services' Secretary, who can subsequently add fiscal and political pressure to states to implement those recommendations. The AC initially examined the recommendations from the report from the American College of Medical Genetics Newborn Screening Expert Panel developed for the Health Resources and Services Administration (HRSA). This panel put forth an initial set of recommendations for inclusion of 29 conditions (and 25 secondary conditions detected in the process of screening for the core 29) for universal screening in 2005.¹ After a period of reflection on this initial work,²⁻⁵ the AC reviewed its procedures for gathering evidence regarding conditions proposed for screening and considered several options for evidence reviews.⁶ Among the guiding principles in this planning were (1) adaptation of established evidence review processes for screening or treatment programs; (2) transparency in the data abstraction and review; (3) recognition of the special challenges regarding evidence about the rare diseases coming to AC consideration; and (4) public access and input to the process.

Several groups and agencies, such as the US Preventive Services Task Force, Evidence-Based Practice Centers sponsored by the Agency for Healthcare Research and Quality, and the Centers for Disease Control and Prevention's (CDC) Task Force on Community Preventive Services, perform systematic

reviews of evidence in areas of health care. The CDC program of Evaluation of Genomic Applications in Practice and Prevention represents another important effort in developing evidence, mainly in the assessment of genetic tests that may have applicability in clinical or public health practice.⁷ Although the US Preventive Services Task Force, in particular, emphasizes studies of population-based screening,^{8,9} most clinical review groups focus more on treatment or clinically based screening. Few, if any, review groups are experienced in evaluations of screening for rare conditions, where evidence from randomized controlled screening or treatment trials is nearly always lacking.

Evidence review groups (ERG) typically begin with a clear case definition of the condition under review, applicable to the subject cohort, to help determine potential benefits of screening or treatment. Review of available newborn screening literature uncovers a complex spectrum of conditions, some of which may lack a clear case definition or appear as a variant along a spectrum of disease severity. Moving from clinical samples to population-based screening and testing typically reveals a broader spectrum. Other constraints on available data for rare diseases include limitations in knowledge of the full impact of disease and the clinical descriptions often limited to case reports. Thus, the cohort of children that clinically presents with the most severe form may overly influence the understanding of key clinical characteristics, natural history, and response to treatment.¹⁰

As with any other set of conditions, these rare diseases too may have different manifestations, courses, and response to treatment arising from a number of other parameters, including differing genetic variants, gender, race/ethnicity, coexisting conditions, treatment centers, or environmental exposures, and sample sizes will in most cases be too small to account for any of these differences. The small number of cases and their clinical progression (i.e., usually severe and often fatal) make it unlikely that investigators will conduct randomized controlled trials in the face of a potentially life-saving or life-changing treatment. The general paucity of data generated by uncontrolled clinical trials can also mean that the full impact of early diagnosis and treatment versus later clinical diagnosis may not be fully known. In the context of screening, the rarity of conditions leads to low true-positive rates (and, thus, much risk of false positives), and studies of screening often require several years even in large populations to document sensitivity and specificity well. Despite its importance for policy making, evidence regarding these conditions typically lacks much information on costs and benefits across all potential outcomes (i.e., true and false positives and negatives), and the available information is not standardized.

Critical sources of information for these rare conditions may also be unpublished—either in hands of investigators who have not yet published the data or, in some cases, with pharmaceutical manufacturers who have tested treatments for rare condi-

From the ¹Center for Child and Adolescent Health Policy, MassGeneral Hospital for Children, Boston, Massachusetts; ²Department of Pediatrics, Harvard Medical School, Boston, Massachusetts; ³Center for Human Genetic Research, MassGeneral Hospital for Children, Boston, Massachusetts; ⁴New England Newborn Screening Program and Department of Pediatrics, University of Massachusetts Medical School, Worcester, Massachusetts; ⁵Department of Pediatrics, Columbia University, New York, New York; ⁶Department of Pediatrics, Child Health Evaluation and Research Unit, University of Michigan Health System, Ann Arbor, Michigan; ⁷New England Connection for PKU and Allied Disorders, Woburn, Massachusetts; and ⁸Duke Clinical Research Institute and Department of Pediatrics, Duke University, Durham, North Carolina.

James M. Perrin, MD, MGH Center for Child and Adolescent Health Policy, 50 Staniford Street, 901, Boston, MA 02114. E-mail: jperrin@partners.org.

Disclosure: The authors declare no conflicts of interest.

Submitted for publication October 30, 2009.

Accepted for publication January 5, 2010.

Published online ahead of print February 11, 2010.

DOI: 10.1097/GIM.0b013e3181d28eb1

tions but not published them. Some data regarding treatments will exist in the Food and Drug Administration trials data, but such data are not available to the general public or for this committee's work.

THE EVIDENCE REVIEW GROUP: PROCESSES AND PROCEDURES

In 2007, the HRSA's Maternal and Child Health Bureau (MCHB), which staffs the AC, entered into an agreement with the MassGeneral Hospital for Children Center for Child and Adolescent Health Policy to outline and initially test a process for systematic evidence development that could support the AC activities.¹¹ The Duke Clinical Research Institute also participates in this effort. In 2008, the HRSA/MCHB, under the AC's authorizing legislation, expanded this activity with MassGeneral to include the promulgation of specific evidence reviews by the Center staff, with the main purpose of providing timely information to the AC to help inform their decision making regarding new conditions nominated for addition to the uniform screening panel. Of note, this review effort provides data to the Bureau but avoids making any recommendations regarding newborn screening. The responsibility for recommendations lies with the AC.

The Center developed a diverse staff and team to perform the reviews, with representation from metabolic genetics, clinical epidemiology, health policy, newborn screening, family advocacy groups, and economics (all authors of this article). This ERG has also developed a team of external consultants, who provide additional and separate review of ERG work before its submission to the Bureau. Especially, given the sensitivity of the questions addressed by the ERG, the Bureau and the ERG considered it critical to develop clear documentation of any conflicts of interest and to exclude anyone with substantive conflicts from participation in the day-to-day workings of the ERG in its development of evidence reports. In cases where a member of the ERG has a direct conflict related to a condition under review, then that person is excused from all review and discussion of the condition. The conflict of interest methods are modeled after those used by the Institute of Medicine and examine intellectual, financial, and other personal and institutional conflicts.

Systematic reviews

The ERG performs systematic reviews of published literature as a key component of its tasks. Each review uses a common base set of questions developed originally through discussions with the AC. In addition, the ERG develops a set of questions specific to the condition under review. The questions usually fall into five main categories: (1) information about the condition itself, (2) evidence regarding screening and screening tests, (3) evidence regarding diagnostic methods, (4) evidence regarding treatment, and (5) economic evaluation. Questions about the condition include whether it is well defined, its prevalence and incidence (including key clinical variations), and its natural history (including the spectrum of severity and variations by key phenotypic or genotypic characteristics). Questions regarding screening include the methods available to screen newborns for the condition; their accuracy; their ability to distinguish early versus late onset cases; their sensitivity, specificity, and predictive values; analytic and clinical validity; and the feasibility of implementing these methods for universal screening.^{12,13} What are the potential harms or risks of screening? What is known about costs and cost-effectiveness of screening? What pilot testing has taken place in population studies or clinical groups?

The main questions about diagnosis address methods and costs of diagnostic testing (similar to those about the screening test) and availability and capability of diagnostic centers. For treatment, does presymptomatic or early symptomatic treatment improve health outcomes and, if so, more than treatment after symptoms develop? What is known about the efficacy and effectiveness of treatment? What is the relationship between treatment timing and treatment outcomes? What are the potential harms or risks of treatment? Is treatment standardized, and where appropriate, is it Food and Drug Administration approved? The economic evaluation questions address the costs and cost-effectiveness of screening. What incremental costs are associated with the use of the screening test in (state) newborn screening programs? What are the costs of diagnosis and the failure to diagnose in the presymptomatic period? What is the availability of treatment and the costs associated with treatment?

Systematic reviews have sought to gather published data for all of these questions. Nonetheless, only a small number of research publications address most of the questions. Although review manuscripts may provide some consensus among investigators or present the studied opinions of well-established investigators, the systematic reviews focus on published research studies rather than consensus or opinions. In general, cost data and especially cost-effectiveness data are rarely available for any of these conditions and represent an area in particular need of dedicated investigation. One issue that arises in reviews is the difficulty in determining whether study cohorts overlap or represent entirely new cases. Although the reviews make a consistent effort to address the full set of questions, key questions that have (so far) driven discussion within the AC have tended to be (a) what is the evidence regarding the feasibility and results of population-based screening; (b) what evidence supports the earlier identification of children with the condition, i.e., will earlier identification improve treatment outcomes; and (c) how effective is treatment?

The ERG has set several parameters for the literature search, usually including limitations to peer reviewed human studies from the past 20 years and only in English. Case studies with fewer than five subjects are excluded. The review provides manuscript details of all excluded single- and small-number case studies in an appendix. As noted above, although review articles and consensus statements are not examined in depth, references in these articles are reviewed to assure inclusion of all relevant research articles. Where a review provides new findings along with sufficient detail regarding methods, we include that material as evidence. Finally, the nomination form for each condition may include up to 15 research articles to support the nomination. The ERG includes articles from the nomination form in the review when they satisfy literature search parameters (i.e., excluding animal studies or review articles).

For each condition, the ERG assigns 2–3 members of the group (core team) to perform the in-depth review, with other members of the group serving as support and commentators on the review progress and findings. The core team then reviews abstracts from all articles returned in the initial search and determines which articles to include in the in-depth literature review, using the same inclusion-exclusion criteria noted above. The next step in the review includes abstraction of all selected articles using standardized tools and assessing the quality of the studies in two different ways. The first is for the quality of the study design¹⁴ (standard form available from the authors) adapted to collect information on the specific questions for the condition. This part of the quality assessment is fairly analogous

to quality assessment used for other evidence review processes, specifically assessing the quality of the study within design categories. For example, a list of criteria assesses cohort studies, and a different list is used to assess case series. The second quality assessment in the abstraction process relates to study goals, adapted from Pandor et al.¹⁵ and Pollitt et al.¹⁶ For example, the type of evidence appropriate for a treatment study differs from evidence for assessing the natural history of a disorder. The core team members all independently review a subsample of 10–20% of all articles to determine reliability of the abstraction process. The core team and staff then collate abstracted data around the key study questions and develop evidence tables from the abstracted articles.

Advocates and experts

The ERG seeks involvement of parent/advocacy groups in its work. Staff, working with the family advocacy consultant to the ERG, and the AC and other consultants identify main advocacy groups interested in the study condition. The ERG team then performs interviews with representatives of those groups, mainly to solicit their perspectives on the key issues regarding the condition, their view of testing and treatment, and other advice they have regarding development of evidence for the AC review.

The ERG recognizes that in a rapidly developing field such as newborn screening, there may be important but unpublished data. The ERG identifies experts, including clinical researchers and newborn screening experts, to help with the identification of this information. These individuals are identified as authors of key articles included in the literature review, through ongoing discussions with content experts and through recommendations from the AC and ERG members.

ERG staff contact experts via e-mail explaining the purpose of the review and sending a conflict of interest form and an open-ended survey including questions from the review. Any expert must provide a conflict of interest form (the same as used by project staff), and this form must be reviewed and in project files before the inclusion of any information from that expert in the ERG report. Conflicts do not exclude experts from providing information on request from the ERG but are maintained by the ERG. The involvement of experts is limited to providing data (clinical and research) regarding the key questions underlying the review, i.e., they are not asked for opinions regarding the evidence or its interpretation, just for additional evidence that may help to inform the AC. Insofar as this information has not undergone peer review, we are unable to apply similar quality assessments as are used with published data. As part of the report preparation, experts are asked to review any material that the ERG reports from their work but not for any interpretation of those data or any recommendations arising from their data or for the AC in other matters regarding the condition under review.

All data and information in the final evidence report become public record. Experts are advised of the public nature of any evidence included in an ERG report. Journal editors have advised that summary evidence presented in these reports should not preclude publication of the full report in a peer reviewed journal, but experts may choose not to provide information to the ERG because of concerns about premature disclosure. For investigators, the balance may be between wanting to provide the AC with the most up-to-date information while preserving some confidentiality before peer reviewed publication. In general, investigators have been generous with their information, given the parameters of the ERG work and its reports. This expert process has provided substantial additional data and clarified key issues in

the ERG reports, in such areas as screening experience and long-term follow-up of treated children.

PRODUCTS

The report developed by the ERG follows the main study questions, highlights the key findings, and indicates where key data are missing. As noted above, the report makes no recommendations (other than for needed research to inform the AC better) nor does it provide any interpretation of the data other than clarifications of the available information. The report, both in preliminary and penultimate versions, undergoes review by the ERG external consultant group. The AC may review preliminary versions of the evidence report and suggest additional questions or data for the ERG to address. Based on these reviews, the ERG prepares a final report, which is presented to the Bureau staff for distribution to the AC. In addition, the ERG formally presents the report at a meeting of the AC.

In its first 15 months, the ERG has performed three in-depth reviews based on these investigative strategies. These reviews have covered Pompe disease, severe combined immunodeficiency, and Krabbe disease. Final reports in each case are available online at the AC Web site.¹⁷

SUMMARY

This article describes the background, development, and initial implementation of new procedures for the systematic review of key issues in newborn screening. Building on the work of other systematic review efforts, the ERG described here has aimed to develop consistent and transparent strategies for evidence review. This process has helped to strengthen a complex analysis and decision system by providing balanced evidence, taking into account available high-quality data, expert opinion, and other levels of evidence, in a transparent manner. The methods developed and the identification of areas of missing data may also help investigators begin to standardize the clinical and laboratory data they collect pertaining to the newborn screening and diagnosis of rare disorders and their outcomes and focus future research efforts in the most needed areas.

ACKNOWLEDGEMENTS

This review was made possible by subcontract number SC-07-028 to Massachusetts General Hospital, Center for Child and Adolescent Health Policy under prime contract number HHSP23320045014XI to Altarm Institute, from the Maternal and Child Health Bureau (MCHB) (Title V, Social Security Act), Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS).

REFERENCES

1. Watson MS, Mann MY, Lloyd-Puryear MA, Rinaldo P, Howell RR, editors. Newborn screening: toward a uniform screening panel and system. *Genet Med* 2006;8(suppl 1):1S–252S.
2. Moyer VA, Calonge N, Teutsch SM, Botkin JR. Expanding newborn screening: process, policy, and priorities. *Hastings Cent Rep* 2008;38:32–39.
3. Howell RR. We need expanded newborn screening. *Pediatrics* 2006;117:1800–1805.
4. Botkin JR, Clayton EW, Fost NC, et al. Newborn screening technology: proceed with caution. *Pediatrics* 2006;117:1793–1799.
5. Lloyd-Puryear MA, Tonniges T, Van Dyck PC, et al. American Academy of Pediatrics Newborn Screening Task Force Recommendations: how far have we come? *Pediatrics* 2006;117:S194–S211.
6. Green NS, Rinaldo P, Brower A, et al; Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children. Committee report:

- advancing the current recommended panel of conditions for newborn screening. *Genet Med* 2007;9:792–796.
7. Available at: <http://www.cdc.gov/genomics/gtesting/EGAPP/index.htm>. Accessed January 9, 2009.
 8. Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20:21–35.
 9. Guirguis-Blake J, Calonge N, Miller T, Siu A, Teutsch S, Whitlock E; U.S. Preventive Services Task Force. Current processes of the US Preventive Services Task Force: refining evidence-based recommendation development. *Ann Intern Med* 2007;147:117–122.
 10. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–930.
 11. See Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, minutes of meeting 10, “Process for nominating and evaluating conditions for inclusion on the uniform newborn screening panel.” Available at: <http://www.hrsa.gov/heritabledisorderscommittee/meetings/2007may/minutes0507.htm>. Accessed August 1, 2009.
 12. Rinaldo P, Zafari 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. Pass K, Green N, Lorey F, Sherwin J, Comeau AM. Pilot programs in newborn screening. *Ment Retard Dev Disabil Res Rev* 2006;12:4:293–300.
 14. NHS Center for Reviews and Dissemination March 2001. Accessed October 17, 2008.
 15. Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S. Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review. *Health Technol Assess* 2004;8:12.
 16. Pollitt RJ, Green A, McCabe CJ, et al. Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assess* 1997;1:i–iv, 1–202.
 17. See Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, Committee Reports. Available at: <http://www.hrsa.gov/heritabledisorderscommittee/reports/>, listed as *KrabbeEvidenceReview FinalDraft.pdf*, *evidencereviewSCID/default.htm*, *evidenceReviewPompe Oct2008.htm*. Accessed October 1, 2009.