

Expanding Newborn Screening:

Process, Policy, and Priorities

by VIRGINIA A. MOYER, NED CALONGE, STEVEN M. TEUTSCH,
AND JEFFREY R. BOTKIN, ON BEHALF OF THE UNITED STATES
PREVENTIVE SERVICES TASK FORCE

In the 1960s, newborn screening programs tested for a single very rare but serious disorder. In recent years, thanks to the development of new screening technology, they have expanded into panels of tests; a federally sponsored expert group has recommended that states test for twenty-nine core disorders and twenty-five secondary disorders. By the standards used to decide whether to introduce new preventive health services into clinical use, the decision-making in newborn screening policy has been lax.

When a new medical technology is implemented in a program that is supported by public funds and mandatory for all children, the evidence that the technology is effective should be solid. Unfortunately, in the last few years, newborn screening policy has not fully followed this principle. More recently, efforts have been made to adhere more closely to an evidence-based approach to newborn screening. In this article we review what has been done in the past in order to inform the process going forward.

All U.S. states and territories sponsor public health programs to screen newborns for selected

hereditary and congenital conditions. The paradigm condition for newborn screening is phenylketonuria, a genetic metabolic disorder that causes permanent mental retardation unless infants who have it are identified before it is clinically apparent and placed on a special diet. State-mandated screening for PKU began in the 1960s and 1970s. Since then, states have added a range of other conditions to their screening panels depending on efficacy, program support, and local advocacy. Because of the paucity of scientific information about efficacy, policy about which tests to make mandatory has historically varied considerably from state to state.

In 2000, a national task force sponsored by the American Academy of Pediatrics and the federal Health Resources and Services Administration called for a standard list of conditions to be developed for the state panels.¹ HRSA then funded the American

Virginia A. Moyer, Ned Calonge, Steven M. Teutsch, and Jeffrey R. Botkin, on behalf of the United States Preventive Services Task Force, "Expanding Newborn Screening: Process, Policy, and Priorities," *Hastings Center Report* 38, no. 3 (2008): 32-39.

College of Medical Genetics to convene a group of newborn screening experts who could fashion recommendations for the list.² In a report released in 2005, the ACMG group called for all states to adopt a core panel consisting of twenty-nine primary disorders for which evidence of benefit was regarded as compelling, as well as twenty-five secondary disorders that would be detected incidentally while screening for the core disorders. The report has been endorsed by an assortment of organizations, including advocacy groups, professional associations, and the Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, an official advisory body to the Secretary of the Department of Health and Human Services.

In this paper, we express serious reservations about the rapid expansion of newborn screening programs that is currently taking place without full consideration of all the issues at stake. Screening recommendations for public health programs should be transparent, unbiased, evidence-based, and attentive to important social values, especially if they will affect every child born in the United States. The original ACMG process did not conform to contemporary standards of evidence-based decision-making. In our view, state and federal policymakers should further evaluate each condition proposed for screening before recommending that it be included in a mandated screening panel.

The Technological Imperative

Except for a test for hearing impairment, all of the tests in the ACMG's uniform panel are blood analyses. The newborn's heel is stuck, blood is collected on an absorbent card, the card is sent to a lab, and the screening tests are carried out. Babies who test positive for a condition are referred for follow-up testing to confirm the diagnosis, since false positives are always possible. If the diagnosis is confirmed, the child must then be linked to appropriate long-

term treatment and management. The test is typically designed to minimize the number of false negatives—that is, results that incorrectly indicate a child does not have a condition—but usually the design involves a tradeoff: changing test cutoff levels to reduce the number of false negatives increases the number of false positives.

The methodology used to evaluate whether to include a condition on a panel placed considerable weight on the multiplex capability of tandem mass spectrometry, giving preference to conditions it detects.

A newborn screening program is not just a panel of screening tests, however. Ideally, it is also parental education, follow-up, diagnosis, treatment and management, and program evaluation, and all of the various parts of the system must be in place and working well to realize the benefits of screening.

Nonetheless, the test itself is what gets the most attention, and the technology available for the test has been a critical factor driving the growth of newborn screening. The invention of the original PKU test and the card used to store and transport the blood sample to the laboratory made public newborn screening programs possible. What has enabled the sudden expansion of newborn screening is the availability of a new technology for conducting the tests. In the past, adding a new condition to a program meant adding new laboratory equipment to an already complex system. Now, however, most of the new conditions recommended by the ACMG group are detected using a single method—tandem mass spectrometry (also known as MS/MS), which produces results with a high degree of precision and accuracy and permits multiplex testing, in which a single

blood analysis screens for many conditions at once.³ Tandem mass spectrometry can replace the tests formerly used to screen for PKU while simultaneously screening for many other metabolic abnormalities—some clinically significant and treatable, some clinically significant but not treatable at this time, and others of unknown significance.

The availability of tandem mass spectrometry has led to considerable pressure to expand newborn screening. Parents of children who have candidate disorders, health professionals who treat these disorders, and private firms that sell screening tests and equipment have all advocated for the expansion of state programs.

Campaigns inspired by the deaths of children with undiagnosed medium chain acyl-coenzyme A dehydrogenase deficiency (MCADD), a disorder of fatty acid metabolism, have been particularly influential. People with MCADD cannot go without food for very long; fasting may cause them to suddenly experience hypoglycemia, vomiting, lethargy, seizures, encephalopathy, coma, apnea, respiratory arrest, cardiac arrest, and sudden death.⁴ Management includes avoiding fasting and taking nutritional supplements. MCADD can be detected through tandem mass spectrometry, and plainly the benefits for some children will be considerable. Advocates have argued very effectively for investing in tandem mass spectrometry to identify children with MCADD, and once the technology is in place, they argue that the state might as well test for the whole range of conditions that

tandem mass spectrometry can detect.

The ACMG's recommendations were influenced by this perspective. The methodology that the ACMG used to evaluate whether to include a condition on a panel placed considerable weight on the multiplex capability of tandem mass spectrometry, giving preference to conditions detectable by the technology and making it more likely that they would be selected for the uniform panel.

However, many of these disorders are poorly understood or not treatable (or both), and screening for such disorders on a population basis departs from standard public health practice. Moreover, a newborn screening panel should be expanded only if the newborn screening program is fully prepared to make all the components of the complex system available for the new disorders. Expansion would be costly and might not be the best use of scarce health care resources, given the many other unmet child health needs. We present the rationale for this perspective in the following sections.

A Conceptual Framework for Analysis of NBS Expansion

Recently, in an important paper in *Health Affairs*, David Atkins and colleagues suggested a framework that policy-makers could use to sort through a controversy about whether to adopt a new technology.⁵ The Atkins framework asks six major questions:

- 1) What is the ultimate goal, and how does the intervention achieve those ends?
- 2) How good is the evidence that the intervention can improve important outcomes?
- 3) How good is the evidence that the intervention will work in the setting specific to the policy-maker?

4) How do the potential benefits compare with the possible harms or costs of the intervention?

5) What constitutes "good enough" evidence for a policy decision?

6) What other considerations are relevant to policy decisions?

The framework is helpful for thinking about the ACMG group's process, its uniform panel recommendations, and the current newborn screening expansion.

The Ultimate Goal

The intervention at issue is mandatory screening of all newborns for certain rare hereditary and congenital disorders, with follow-up to confirm diagnosis and initiate treatment or management of the condition. Justifying this intervention requires that we make a special case for its mandatory nature. Medical screening of children for a health condition normally requires parental informed consent, especially when the condition is not a threat to others in the community.

Several different rationales have been suggested for newborn screening, and whether screening can justifiably be mandatory may be considered independently for each. The most important and widely accepted goal of newborn screening is to improve health outcomes in the screened population of newborns. Given this goal, screening makes sense only if early detection and treatment will lead to better health outcomes than would be possible if treatment were delayed until the condition became symptomatic.

Meeting this condition requires that there be both a suitable test and a treatment that is effective and works better if delivered before symptoms appear. When PKU screening was introduced, for example, proponents argued that mandatory, universal screening was necessary because the

consequences of untreated PKU were so dire and the treatment so straightforward and effective. Currently, newborns are screened for PKU without parental informed consent in all but a few states. (Some states give parents the right to refuse PKU screening, although parents frequently do not understand that they have that right.)

While not everyone accepts the urgency argument for foregoing informed consent, even for PKU, others have argued for less stringent rationales for newborn screening. For example, they assert that mandatory screening can also be justified if, by identifying infants with rare conditions, it facilitates the research necessary to develop effective treatments.

We find this rationale to be ethically questionable. Mandating screening in order to recruit human research subjects does not conform to standard ethical or privacy requirements. Further, research on rare metabolic conditions cannot generate useful information without a research infrastructure that supports collaborative clinical trials. Such an infrastructure exists for childhood cancer, but not yet for rare metabolic conditions.

Some argue that screening to identify potential subjects may benefit the infants.⁶ In pediatric cancer care, for example, access to clinical trials enhances the overall quality of care, especially for those children with poor prognoses. However, there is no mandated testing of asymptomatic children for cancer in order to find children to enroll in clinical trials of cancer treatment. We think that, at the very least, informed decision-making by the parents should be required prior to screening if the primary goal is to identify potential subjects for research.

Another goal sometimes suggested for newborn screening is the provision of information to parents about a child's health status. Advocates point to several reasons that providing information is an appropriate screening goal even when a condition has no proven medical treatment:

- When the child develops symptoms, the family can avoid the so-called diagnostic odyssey—the protracted search for an explanation of a health problem.
- The parents can avoid the impact of a second affected child on the family by incorporating information about an inherited condition into their future reproductive decisions.⁷
- If the parents know something about a child’s future health problems, they can make plans for managing the impact of the condition on the child and the family.
- Knowing the information is valuable in itself.

We find these points insufficient to justify mandated public health screening of all newborns. First, a diagnostic odyssey does not begin until a condition becomes clinically manifest. If treatment need not begin until after symptoms appear, then why mandate screening in newborns? Testing individuals might be more effective when symptoms first appear, since it would lead to fewer false positives. Clinical strategies to improve diagnosis should also be considered. Better clinical strategies would reduce the probability of diagnostic odysseys and are important even if newborns are screened. Since screening always produces some false negatives, physicians must be able to recognize the clinical presentation of a condition regardless of whether it is included in a mandatory screening panel.

On the second point, opinions will differ. Some parents may welcome information about the risk of an inherited disorder in a future pregnancy, but others may not. In this country, there is a strong ethical presumption that adults should decide what genetic information they wish to have about themselves, and there is an even stronger presumption that they should make their own reproductive decisions. If the provision of informa-

tion for reproductive decisions is the goal of screening, then parents should give their informed consent to the screening.

Similarly, while some parent advocates speak eloquently about their desire to know about a child’s condition at birth so they can prepare themselves to provide appropriate care, others may prefer to remain ignorant until symptoms appear if the child will realize no benefit from treatment administered before then. Given the wide range in clinical presentations

Several different rationales have been suggested for mandatory newborn screening. Some assert that it can also be justified if, by identifying infants with rare conditions, screening facilitates research on treatments.

and the consequent uncertainty about how an individual child will be affected, the latter preference is quite reasonable, and obtaining parental informed consent would respect it.

Finally, some have argued that families have a “right to know” about genetic diseases in their children even if no effective treatment is available.⁸ In particular, some argue that if tandem mass spectrometry is used at all, it should be used to test for all the abnormalities it is capable of detecting, and the information should be provided to parents. This goal seems far removed from the goal of improving health outcomes, however, and it violates a time-honored tenet of medicine that clinicians should not order a test if the results will not change clinical management. Moreover, providing the information may have bad consequences for the children and their families—*anxiety, changes in family relationships and dynamics, unnecessary treatment, and labeling that could lead to uninsurability.*

Certainly some parents will want to obtain information about a child’s

health status. If mandated public health screening of newborns is justified on other grounds, then the information it produces might reasonably be an additional benefit, at least for some parents, as long as the harms are also taken into account. We believe, however, that the goals of identifying potential research subjects and providing parents with information about a child’s future health status do not in themselves justify mandated screening of all newborns. Detecting disorders that have no proven treat-

ment or for which treatment is helpful only after clinical presentation is just not as urgent as detecting PKU.⁹

Evidence of Effectiveness

In order to assess whether newborn screening can improve health outcomes as expected, policy-makers need a great deal of information about the conditions, the associated screening tests, and the entire newborn screening program structure. For each condition, they need information on incidence and natural history. How many infants will be identified as having the condition, and of those identified, how many will go on to develop noticeable symptoms? Even when the natural history of clinically detected cases of a condition is known, the natural history for *screen-detected* cases is often poorly understood. Many children with screen-detected conditions may never develop clinically important morbidity and mortality, or they may be likelier to have milder cases than those who are detected clinically.

Policy-makers also need information on the availability of effective treatments and whether early detection provides enough advantages to warrant screening. To justify mandatory public health screening of all newborns, proven effective therapies or preventive strategies for included diseases should be available, and they should be more effective when provided before symptoms appear. Information on the risks of treatment is also important. Unproven and potentially harmful treatments are particularly bad for children who would not

highly effective in reducing morbidity and mortality from the disease. Several studies have found, however, that fewer than half of affected children receive the recommended prophylactic measures, such as antibiotics and immunizations.¹¹

How good is the evidence of improvement and effectiveness for the disorders on the panel that the ACMG group has recommended? It's hard to say. The push for expanded newborn screening has bypassed traditional, evidence-based decision-making processes at both the state

and federal levels. The ACMG group's task was to review the evidence on the various conditions and decide which should be included in the uniform panel. However, it developed its own process—one that was neither transparent nor open to independent review. No experts in systematic reviews or evidence-based recommendation development were invited to participate or comment. The United States Preventive Services Task Force, the Evidence-Based Practice Centers sponsored by the Agency for Healthcare Research and Quality, the Centers for Disease Control's Task Force on Community Preventive Services, and the Institute of Medicine were given no role in the process. Other than specific newborn screening specialists, state policy-makers were not involved, even though most states have newborn screening advisory committees to advise the executive and legislative branches of state government, and two states (Massachusetts and Washington) conduct structured reviews of evidence pertaining to screening tests. The process and

the grounds for the ACMG group's recommendations were eventually outlined in a lengthy report, and HRSA announced a sixty-day public comment period, but the report was released for review and comment months after the recommendations had become public knowledge and had been endorsed and promoted by advocacy groups. The detailed report describes a flawed process. In determining the recommended panel of tests, the ACMG report relied on an opinion survey chiefly of disease experts, screening specialists, and lay and professional advocates for screening, supplemented by unsystematic reviews conducted by selected disease experts. The ACMG group developed its own criteria and weighting system to prioritize disorders for inclusion. The system appeared to give as much or more weight to the testing technology as to the health benefits of early detection, with a strong preference for the capability of detecting multiple disorders. Supplemental information about the estimated prevalence of these disorders demonstrated the lack of robust epidemiological data. Very little attention was given to concerns about the quality of the evidence, the costs of expansion, or potential harms from false positive screening results or potentially unnecessary treatments.

The report describes the twenty-nine disorders on the recommended core panel as meeting three "minimum" criteria: each condition is identifiable twenty-four to forty-eight hours after birth; a high-throughput screening test with "appropriate sensitivity and specificity" is available; and there are "demonstrated benefits of early detection, timely intervention, and efficacious treatment." The first criterion is straightforward. The second is obscure, since "appropriate" is not explained. Moreover, for many of the conditions, there has been little or no recorded experience with large-scale population screening. Thus, what it means to meet this criterion is hard to say.

The incidence and natural history of most abnormalities identified through tandem mass spectrometry is uncertain, which means there is uncertainty about the consequences of treatment.

have progressed to clinically important disease in the first place.

Policy-makers also need information on the characteristics of the screening tests, such as rates of false positives and false negatives. When the entire newborn population is tested, even low rates of false positives with a very accurate test will lead to further testing for many babies. Whether a large number of false positives is acceptable depends on the extent to which morbidity and mortality are prevented or ameliorated by screening.

Finally, in order to ensure effectiveness in local settings, policy-makers must have information about a state's ability to put all the pieces in place and create a system that works for each disorder over the entire range of circumstances prevailing throughout the state. Even when a screening intervention is shown to be efficacious under controlled or experimental circumstances, in actual practice it may be substantially less.¹⁰ For example, early detection and prophylactic treatment for sickle cell disease are

and federal levels. The ACMG group's task was to review the evidence on the various conditions and decide which should be included in the uniform panel. However, it developed its own process—one that was neither transparent nor open to independent review. No experts in systematic reviews or evidence-based recommendation development were invited to participate or comment. The United States Preventive Services Task Force, the Evidence-Based Practice Centers sponsored by the Agency for Healthcare Research and Quality, the Centers for Disease Control's Task Force on Community Preventive Services, and the Institute of Medicine were given no role in the process. Other than specific newborn screening specialists, state policy-makers were not involved, even though most states have newborn screening advisory committees to advise the executive and legislative branches of state government, and two states (Massachusetts and Washington) conduct structured reviews of evidence pertaining to screening tests. The process and

The third criterion is the most important, but based on the information provided in the report, the unequivocal statement that all twenty-nine core disorders satisfy it is also difficult to assess. For the group of metabolic abnormalities identified through tandem mass spectrometry testing, some conditions are good candidates for early detection and screening, given the evidence available. For example, good cases can be made for early identification and management of PKU and MCADD. There is substantial experience with population-based screening and treatment of PKU and fairly good information about the prevalence of MCADD, as well as some information about its natural history. Also, intervention appears to reduce the risk of fatal metabolic crisis and has few anticipated harms (although the latter have not yet been well characterized).¹²

For most other abnormalities identified through tandem mass spectrometry, logic and a close reading of the report reveal much greater uncertainty about the incidence of these abnormalities and their natural history in individuals identified through screening. Given this fact, there is inevitably less direct evidence of benefit and more uncertainty about the health consequences of treatment. Left unscreened, some children might never have known about their conditions because they might never have developed symptoms. Also, studies of some conditions identify variants that do not affect the patient's health, and this variation argues against routinely screening for such conditions.¹³ Even with MCADD, population screening studies suggest there are less severe genetic variants for which the impact of early detection and treatment is uncertain.¹⁴

In a document titled, "Conceptualizing and Combining Evidence for Health System Guidance," the Canadian Health Services Research Foundation identifies three types of evidence that evidence-based decision-making may draw on: colloquial evi-

dence, context-free scientific evidence, and contextual scientific evidence.¹⁵ Colloquial evidence includes evidence about resources, expert opinion, political judgment, values, habits, traditions, special interests, and other elements of the specific issue. Context-free scientific evidence consists of truths that are valid in any context; this is what evidence-based medicine depends on most heavily. Contextual scientific evidence consists of truths that are dependent on the characteristics of the setting in which the intervention takes place.

All of the conditions on the ACMG core list can be supported by colloquial evidence, but only a small number are well supported by either type of scientific evidence. The opinion survey of disease experts and others was (at best) colloquial evidence. In the second stage of the process, the evidence base was assessed and a fact sheet was prepared for each disorder. At least two recognized experts on each disorder validated the fact sheets. In the process, they ranked the available evidence about the condition, the test, the diagnosis, and the treatment on a scale of one to four, with one being the gold standard: level one evidence is derived from well-designed randomized controlled trials or diagnostic studies on relevant populations.¹⁶ (The rankings for each disorder are shown in Appendix 1 of the report.) For many conditions, at least one of the two experts categorized the available evidence as level three or four.¹⁷ In short, to argue that screening will be good for children's health, the ACMG group (and other advocates of screening) have at times given up on good scientific evidence and relied on extrapolation and supposition. (And, as noted, they have also argued that screening can be justified by goals other than direct health benefit to the child.¹⁸)

In addition to assessing the evidence related to specific disorders, the ACMG group assessed the ability of states to carry out the activities required to make screening for the uniform panel effective in local settings.

The report identifies significant barriers to the construction of a model newborn screening system, including inadequate state financing, fragmented service delivery, limited availability of metabolic disorder specialists, and the absence of universal health coverage. Both the ACMG report and the American Academy of Pediatrics' 2000 newborn screening task force report indicate that many states have been struggling to overcome these barriers for the conditions already in their panels; if so, adding new conditions will be difficult. In our judgment, the ACMG group's assessment provides reason for concern about the extent to which its uniform panel can be effectively implemented on a national basis.

Benefits and Harms

In addition to the potential risks of treatment, all screening tests have more general potential harms: they may generate unnecessary worry, lead children to be labeled as having serious health problems, and have long-term consequences for insurability and employability. If there were at least fair evidence that an intervention was effective, the benefits could be large enough and certain enough to allow a rough judgment that they outweighed the potential harms. However, when neither benefits nor harms are well characterized, a more cautious approach is warranted.¹⁹ If the risk-benefit evaluation is inadequate, the program could end up doing more harm than good. It might also simply provide no demonstrable health benefit to the children while spending scarce resources. Given tight state budgets for newborn screening programs and other essential child care services, resource consumption without benefit must itself be considered a potential harm.

Sufficiency of Evidence

The fifth question in the Atkins' framework for deciding whether to adopt a new technology is, What

constitutes “good enough” evidence for a policy decision? The ACMG report appears to set the bar low.

The United States Preventive Services Task Force (on whose behalf we write) also faces this question. The USPSTF makes decisions about whether to introduce preventive services, including screening tests, into routine clinical care. The recommendations of the USPSTF are based on comprehensive, systematic evidence reviews and assessments. Recommendations are made only when there is evidence of at least fair quality that the preventive service will result in real net benefit. In the absence of good or fair quality evidence, the USPSTF issues an “I Statement”: there is insufficient evidence to recommend for or against.

Based on the ACMG fact sheets and the validation reports characterizing the evidence, we believe that if the ACMG list of core conditions were evaluated using the USPSTF approach, a few would be recommended with an A or B grade, meaning that there is at least fair evidence that the benefits outweigh the harms. Perhaps a few more would receive C grades; the evidence is at least fair, but what the evidence shows is that benefits and harms are too closely balanced to support an across-the-board recommendation about introducing the service. The majority, however, would be given an I: there is insufficient evidence to recommend for or against. Without performing full-scale evidence reviews of our own, of course, we cannot be sure; but if we are right, the ACMG recommendation that these conditions be adopted by all state newborn screening programs is premature. Good public health practice and good ethics require that the evidence threshold be at least as high for a recommendation for mandatory screening of all newborns through a public health program as it is for recommendation of a nonmandatory test for use in clinical care with informed consent.

The fact that the ACMG’s expert group was composed primarily of

people in the newborn screening field is also a matter for concern. Current U.S. practice suggests that whether evidence is thought to be sufficient to introduce a new screening test depends heavily on who evaluates it; advocates of screening would likely reach different conclusions than an independent decision-making body that incorporated health policy experts. In many states, colloquial evidence provided by advocates and medical experts has been the dominant influence. Two states that conducted structured reviews of scientific evidence, Massachusetts and Washington, have implemented tandem mass spectrometry programs that mandate screening for far fewer conditions than are called for by the ACMG core panel.²⁰ Massachusetts mandates screening for MCADD but makes other testing optional. Washington currently includes MCADD in its screening panel and is in the process of reconsidering other disorders in the ACMG core panel.

Other Relevant Considerations

The paradigm of evidence-based medicine has recently shifted somewhat to include nonscientific information, such as the values of the individual being tested or treated. At the policy level, for example, Neal Kohatsu has defined evidence-based public health as the integration of science-based interventions with community preferences.²¹ Similarly, the Canadian Health Services Research Foundation states that the three types of evidence it identifies—colloquial evidence, context-free scientific evidence, and context-sensitive scientific evidence—require a deliberative process including consultation with relevant stakeholders in order to reach an evidence-based judgment.

The same is true of the decisions regarding newborn screening: those ultimately responsible for setting health policy at the national, state, or even health care system level must balance the scientific and other evidence supporting screening. Since

collective resources, both public and private, support public newborn screening programs, decision-makers must also consider the preferences of the entire community concerning the use of those resources, not just the preferences of those directly concerned with newborn screening. Finally, they should ensure that their decisions respect societal ethics relating to the nature of benefit and harm and to the significance of consent to treatment and research.

It seems clear that the ACMG’s approach, which relied mostly on colloquial evidence, failed to adequately demonstrate effectiveness. It also failed to incorporate the views of the community at large on the proper place of newborn screening within the allocation of health care resources to child health needs.

Recommendations

A high standard of evidence should be met before requiring that all infants be screened for a disorder. We believe that at least some of the twenty-nine conditions ACMG has identified as “core conditions” for screening do not meet conventional population screening criteria, including the minimal criteria proposed by the ACMG report. We do not know how many of these conditions would meet objective criteria for population screening because the process by which this list was produced excluded the evidence-based approaches accepted by the research community in evaluating medical and public health interventions. Perhaps the rarity of candidate conditions for screening and the desire to support a politically popular program derailed the process.

State policy-makers should ask some probing questions before following the ACMG recommendations. Parents should be able to choose whether their children will undergo screening when the evidence on benefits and harms is equivocal or limited. If policy is to move in that direction, however, more attention must be given to ways of involving

parents as active participants in newborn screening programs, and research is needed on prenatal education and prenatal permission for unproven newborn screening.

We suggest that both state and federal governments should objectively evaluate each condition on the basis of prevention potential, medical rationale, treatment availability, public health rationale, available technology, and cost-effectiveness before recommending inclusion in mandated screening panels. Stakeholders, including content experts and advocacy organizations, should participate in this process, but not to the exclusion of evidence-based policy experts who are experienced in the objective evaluation of scientific evidence. Because the scientific evidence is rapidly evolving, we recommend that states revisit their lists on a regular basis—perhaps every three to five years.

Finally, we urge that if states expand newborn screening to include disorders for which the evidence of benefits and harms is incomplete, then they should commit to collecting longitudinal data on infants who test positive. Although the data will be strongly biased (because all or nearly all children identified by screening are likely to be treated, regardless of whether the treatment will alter health outcomes), this information should help us learn from our experiences and implement truly effective, evidence-based screening programs. Given the imminent arrival of yet newer and more powerful screening technologies, we should put a high priority on getting the processes in place to assure sound, evidence-based decisions.

The Advisory Committee on Heritable Disorders and Genetic Diseases of Newborns and Children has moved forward on many of the issues recommended in this paper. They have committed to directing a systematic evidence review for all new conditions nominated by experts and advocates for consideration as additions to the current testing recommendations, and to use these reviews

to make evidence-based recommendations. This is an important step forward in newborn screening policy that should not only inform recommendations about new tests, but also be used to review and update recommendations about tests on the current core list. We strongly support this initiative and encourage full funding so that the process can proceed without delay.

References

1. American Academy of Pediatrics, Newborn Screening Task Force, "Serving the Family from Birth to the Medical Home: Newborn Screening a Blueprint for the Future," *Pediatrics* 106, no. 2, supplement (2000): 389-427.
2. American College of Medical Genetics Newborn Screening Expert Group, "Newborn Screening: Toward a Uniform Screening Panel and System: Executive Summary and Main Report," *Genetics in Medicine* 9, no. 5, supplement (2006): 1S-252S.
3. D. Marsden, C. Larson, and H.L. Levy, "Newborn Screening for Metabolic Disorders," *Journal of Pediatrics* 148 (2006): 577-84.
4. Centers for Disease Control, National Office of Public Health Genomics, "MCAD Deficiency Medium-chain acyl-CoA dehydrogenase," fact sheet; http://www.cdc.gov/genomics/hugenet/factsheets/FS_.htm, accessed September 11, 2007.
5. D. Atkins, J. Siegel, and J. Slutsky, "Making Policy When the Evidence Is in Dispute," *Health Affairs* 24, no. 1 (2005): 102-113.
6. D. Alexander and P.C. van Dyck, "A Vision of the Future of Newborn Screening," *Pediatrics* 117, no. 5 (2006): S350-54.
7. *Ibid.*
8. *Ibid.*
9. S.D. Grosse et al., "From Public Health Emergency to Public Health Service: The Implications of Evolving Criteria for Newborn Screening Panels," *Pediatrics* 117 (2006): 923-29.
10. J.R. Botkin, "Research for Newborn Screening: Developing a National Framework," *Pediatrics* 116, no. 4 (2005): 862-71.
11. S.J. Teach, K.A. Lillis, and M. Grossi, "Compliance with Penicillin Prophylaxis in Patients with Sickle Cell Disease," *Archives of Pediatric and Adolescent Medicine* 152 (1998): 274-78; "Update: Newborn Screening for Sickle Cell Disease: California, Illinois, and New York, 1998," *Morbidity and*

Mortality Weekly Report 49, no. 32 (2000): 729-31.

12. S.D. Grosse and C. Dezateux, "Newborn Screening for Inherited Metabolic Disease," *Lancet* 369 (2007): 5-6.

13. B.T. van Maldegem et al., "Clinical, Biochemical, and Genetic Heterogeneity in Short-Chain Acyl-Coenzyme A Dehydrogenase Deficiency," *Journal of the American Medical Association* 296 (2006): 943-95.

14. B. Wilcken et al., "Screening Newborns for Inborn Errors of Metabolism by Tandem Mass Spectrometry," *Journal of the American Medical Association* 348 (2003): 2304-2312.

15. J. Lomas et al., for the Canadian Health Services Research Foundation, "Conceptualizing and Combining Evidence for Health Systems Guidance," 2005, http://www.chsr.ca/other_documents/evidence_e.php.

16. The evidence levels used in the American College of Medical Genetics report were determined as follows: Level one evidence was derived from well-designed randomized controlled trials or diagnostic studies on relevant populations. Level two evidence was derived from well-designed randomized controlled trials or diagnostic studies with minor limitations and/or overwhelming, consistent evidence from observational studies. Level three evidence was derived solely from observational studies. Level four evidence was derived from expert opinion, case reports, and reasoning from basic principles.

17. For twelve of the twenty-nine core conditions, the experts' rankings of the evidence levels differed by more than one category; for example, one expert might consider the evidence for the benefit of treatment to be level one, and another might consider it level three. This is surprising, since the definitions of the levels are clear. One would expect more agreement from experts reviewing a single body of literature.

18. R.R. Howell, "We Need Expanded Newborn Screening," *Pediatrics* 117, no. 5 (2006): 1800-1805.

19. J.R. Botkin et al., "Newborn Screening Technology: Proceed With Caution," *Pediatrics* 117, no. 5 (2006): 1793-99.

20. K. Atkinson et al., "A Public Health Response to Emerging Technology: Expansion of the Massachusetts Newborn Screening Program," *Public Health Reports* 116 (2001): 122-31; S.D. Grosse and M. Gwinn, "Assisting States in Assessing Newborn Screening Options," *Public Health Reports* 116 (2001): 169-72.

21. N.D. Kohatsu et al., "Evidence-Based Public Health: An Evolving Concept," *American Journal of Preventive Medicine* 27, no. 5 (2004): 417-21.