

Michael S. Watson, PhD, Michele A. Lloyd-Puryear, MD, PhD, Marie Y. Mann, MD, MPH, Piero Rinaldo, MD, PhD, and R. Rodney Howell, MD, editors

Background: States vary widely in their use of newborn screening tests, with some mandating screening for as few as three conditions and others mandating as many as 43 conditions, including varying numbers of the 40+ conditions that can be detected by tandem mass spectrometry (MS/MS). There has been no national guidance on the best candidate conditions for newborn screening since the National Academy of Sciences report of 1975¹ and the United States Congress Office of Technology Assessment report of 1988,² despite rapid developments since then in genetics, in screening technologies, and in some treatments. **Objectives:** In 2002, the Maternal and Child Health Bureau (MCHB) of the Health Resources and Services Administration (HRSA) of the United States Department of Health and Human Services (DHHS) commissioned the American College of Medical Genetics (ACMG) to:

1. Conduct an analysis of the scientific literature on the effectiveness of newborn screening.
2. Gather expert opinion to delineate the best evidence for screening for specified conditions and develop recommendations focused on newborn screening, including but not limited to the development of a uniform condition panel.
3. Consider other components of the newborn screening system that are critical to achieving the expected outcomes in those screened.

Methods: A group of experts in various areas of subspecialty medicine and primary care, health policy, law, public health, and consumers worked with a steering committee and several expert work groups, using a two-tiered approach to assess and rank conditions. A first step was developing a set of principles to guide the analysis. This was followed by developing criteria by which conditions could be evaluated, and then identifying the conditions to be evaluated. A large and broadly representative group of experts was asked to provide their opinions on the extent to which particular conditions met the selected criteria, relying on supporting evidence and references from the scientific literature. The criteria were distributed among three main categories for each condition:

1. The availability and characteristics of the screening test;
2. The availability and complexity of diagnostic services; and
3. The availability and efficacy of treatments related to the conditions. A survey process utilizing a data collection instrument was used to gather expert opinion on the conditions in the first tier of the assessment. The data collection format and survey provided the opportunity to quantify expert opinion and to obtain the views of a diverse set of interest groups (necessary due to the subjective nature of some of the criteria). Statistical analysis of data produced a score for each condition, which determined its ranking and initial placement in one of three categories (high scoring, moderately scoring, or low scoring/absence of a newborn screening test). In the second tier of these analyses, the evidence base related to each condition was assessed in depth (e.g., via systematic reviews of reference lists including MedLine, PubMed and others; books; Internet searches; professional guidelines; clinical evidence; and cost/economic evidence and modeling). The fact sheets reflecting these analyses were evaluated by at least two acknowledged experts for

¹ A medical food is prescribed by a physician when a patient has special nutrient needs in order to manage a disease or health condition, and the patient is under the physician's ongoing care. The label must clearly state that the product is intended to be used to manage a specific medical disorder or condition. An example of a medical food is a food for use by persons with PKU, i.e., foods formulated to be free of the amino acid phenylalanine.

² The Health Insurance Portability and Accountability Act of 1996 (HIPAA) provides relevant protections regarding patient privacy. The federal privacy regulations do not prohibit or interfere with newborn screening and follow-up. Covered entities must track disclosures made without written patient authorization for services other than treatment, payment, and operations, so that the covered entity can provide accounting on patient request. A discussion of the HIPAA issues relating to newborn screening in the context of public health is available in Appendix 4.

³ This and the following economic analyses may best be done through the funding of special projects due to the expense of documentation

⁴ Consider collecting data from a subset that includes all screen-positive newborns from which an overall rate can be extrapolated with minimal increased cost to the program. Consider initially collecting data from a subset that includes all screen-positive newborns for which the data already is needed. From these, an overall rate can be extrapolated with minimal increased cost. The goal is to know all and is dependant on the development of databases in which this information can be maintained and would be facilitated by inclusion on blood collection cards. Identification of undocumented newborns is increasingly important to their participation in such programs. This is an important issue that involves States, hospitals, providers, insurers, and mothers.

⁵ For a guidance article on the HIPAA Privacy Rule and Public Health written by CDC and DHHS. see the *Morbidity and Mortality Weekly Report* for April 11, 2003, vol. 52 pp. 1-21, and www.cdc.gov/privacypolicies and www.hrsa.gov/website.htm.

DOI: 10.1097/01.gim.0000223467.60151.02

each condition. These experts assessed the data and the associated references related to each criterion and provided corrections where appropriate, assigned a value to the level of evidence and the quality of the studies that established the evidence base, and determined whether there were significant variances from the survey data. Survey results were subsequently realigned with the evidence obtained from the scientific literature during the second-tier analysis for all objective criteria, based on input from at least three acknowledged experts in each condition. The information from these two tiers of assessment was then considered with regard to the overriding principles and other technology or condition-specific recommendations. On the basis of this information, conditions were assigned to one of three categories as described above:

1. Core Panel;
2. Secondary Targets (conditions that are part of the differential diagnosis of a core panel condition.); and
3. Not Appropriate for Newborn Screening (either no newborn screening test is available or there is poor performance with regard to multiple other evaluation criteria).

ACMG also considered features of optimal newborn screening programs beyond the tests themselves by assessing the degree to which programs met certain goals (e.g., availability of educational programs, proportions of newborns screened and followed up). Assessments were based on the input of experts serving in various capacities in newborn screening programs and on 2002 data provided by the programs of the National Newborn Screening and Genetics Resource Center (NNSGRC). In addition, a brief cost-effectiveness assessment of newborn screening was conducted.

Results:

Uniform panel – A total of 292 individuals determined to be generally representative of the regional distribution of the United States population and of areas of expertise or involvement in newborn screening provided a total of 3,949 evaluations of 84 conditions. For each condition, the responses of at least three experts in that condition were compared with those of all respondents for that condition and found to be consistent. A score of 1,200 on the data collection instrument provided a logical separation point between high scoring conditions (1,200– 1,799 of a possible 2,100) and low scoring (<1,000) conditions. A group of conditions with intermediate scores (1,000–1,199) was identified, all of which were part of the differential diagnosis of a high scoring condition or apparent in the result of the multiplex assay. Some are identified by screening laboratories and others by diagnostic laboratories. This group was designated as a “secondary target” category for which the program must report the diagnostic result.

Using the validated evidence base and expert opinion, each condition that had previously been assigned to a category based on scores gathered through the data collection instrument was reconsidered. Again, the factors taken into consideration were: 1) available scientific evidence; 2) availability of a screening test; 3) presence of an efficacious treatment; 4) adequate understanding of the natural history of the condition; and 5) whether the condition was either part of the differential diagnosis of another condition or whether the screening test results related to a clinically significant condition.

The conditions were then assigned to one of three categories as previously described (core panel, secondary targets, or not appropriate for Newborn Screening).

Among the 29 conditions assigned to the core panel are three hemoglobinopathies associated with a Hb/S allele, six amino acidurias, five disorders of fatty oxidation, nine organic acidurias, and six unrelated conditions (congenital hypothyroidism (CH), biotinidase deficiency (BIOT), congenital adrenal hyperplasia (CAH), classical galactosemia (GALT), hearing loss (HEAR) and cystic fibrosis (CF)). Twenty-three of the 29 conditions in the core panel are identified with multiplex technologies such as tandem mass spectrometry (MS/MS) or high pressure liquid chromatography (HPLC). On the basis of the evidence, six of the 35 conditions initially placed in the core panel were moved into the secondary target category, which expanded to 25 conditions. Test results not associated with potential disease in the infant (e.g., carriers) were also placed in the secondary target category. When newborn screening laboratory results definitively establish carrier status, the result should be made available to the health care professional community and families.

Twenty-seven conditions were determined to be inappropriate for newborn screening at this time.

Conditions with limited evidence reported in the scientific literature were more difficult to evaluate, quantify and place in one of the three categories. In addition, many conditions were found to occur in multiple forms distinguished by age-of-onset, severity, or other features. Further, unless a condition was already included in newborn screening programs, there was a potential for bias in the information related to some criteria. In such circumstances, the quality of the studies underlying the data such as expert opinion that considered case reports and reasoning from first principles determined the placement of the conditions into particular categories.

Newborn screening program optimization – Assessment of the activities of newborn screening programs, based on program reports, was done for the six program components: education; screening; follow-up; diagnostic confirmation; management; and program evaluation. Considerable variation was found between programs with regard to whether particular aspects (e.g., prenatal education program availability, tracking of specimen collection and delivery) were included and the degree to which they are provided. Newborn screening program evaluation systems also were assessed in order to determine their adequacy and uniformity with the goal being to improve interprogram evaluation and comparison to ensure that the expected outcomes from having been identified in screening are realized. **Conclusions:** The state of the published evidence in the fast-moving worlds of newborn screening and medical genetics has not kept up with the implementation of new technologies, thus requiring the considerable use of expert opinion to develop recommendations about a core panel of conditions for newborn screening. Twenty-nine conditions were identified as primary targets for screening from which all components of the newborn screening system should be maximized. An additional 25 conditions were listed that could be identified in the course of screening for core panel conditions. Programs are obligated to establish a diagnosis and communicate the result to the health care provider and family. It is recognized that screening may not have been maximized for the detection of these secondary conditions but that some proportion of such cases may be found among those screened for core panel conditions. With additional screening, greater training of primary care health care professionals and subspecialists will be needed, as will the development of an infrastructure for appropriate follow-up and management throughout the lives of children who have been identified as having one of these rare conditions. Recommended actions to overcome barriers to an optimal newborn screening system include:

- The establishment of a national role in the scientific evaluation of conditions and the technologies by which they are screened;
- Standardization of case definitions and reporting procedures;
- Enhanced oversight of hospital-based screening activities;
- Long-term data collection and surveillance; and
- Consideration of the financial needs of programs to allow them to deliver the appropriate services to the screened population. **Genet Med 2006;8(5, Supplement):12S–252S.**

INTRODUCTION

The work reported here is pursuant to the HRSA/MCHB Contract No. 240-01-0038, *Standardization of Outcomes and Guidelines for State Newborn Screening Programs*. In 1999, the American Academy of Pediatrics (AAP) Newborn Screening Task Force recommended that, “HRSA should engage in a national process involving government, professionals, and consumers to advance the recommendations of this Task Force and assist in the development and implementation of nationally recognized newborn screening system standards and policies.” The Task Force was concerned about the lack of unifor-

mity among states, particularly with regard to their newborn screening condition panels.

In 2001, in response to that recommendation, HRSA/MCHB requested that ACMG outline a process of standardization of outcomes and guidelines for State newborn screening programs and define responsibilities for collecting and evaluating outcome data, including a recommended uniform panel of conditions to include in State newborn screening programs. It was expected that the analytical endeavor and subsequent recommendations be definitive and that the recommendations be based on the best scientific evidence and analysis of that evidence. ACMG was specifically asked to develop recommendations to address:

1. A uniform condition panel (including implementation methodology);
2. Model policies and procedures for State newborn screening programs (with consideration of a national model);
3. Model minimum standards for State newborn screening programs (with consideration of national oversight);
4. A model decision matrix for consideration of State newborn screening program expansion; and
5. Consideration of the value of a national process for quality assurance and oversight.

This report is a product of the work undertaken by ACMG for HRSA. A methods section begins by providing the broad context for the newborn screening system and the overarching principles for developing newborn screening guidelines. It then provides the criteria that were used in the analyses of conditions under consideration for newborn screening programs. This is followed by a description of the development and use of tools to collect data that would complement evidence gathered from a review of the scientific literature, and also by a description of the process for obtaining additional expert information and opinion. The results of these analyses are provided, as well as recommendations for moving forward.

Although the criteria by which the conditions are evaluated and the results of those evaluations are the primary goals of this effort, associated and supporting goals also are described because of their relevance to the newborn screening system. In order to realize the expected outcomes for newborns and their families, the full system must be operating efficiently and effectively.³⁻⁶ Efforts have been made to assess the newborn screening system based on its component parts, which allows for the development of specific standards for program performance and for an assessment of status of the programs. This assessment also provides the opportunity to determine the extent to which a systematic national approach to quality assessment and assurance is possible.

SECTION I: DEVELOPING A UNIFORM SCREENING PANEL

A. Background

In the United States, newborn screening is a highly visible and important State-based public health program^{2,7-10} that began over 40 years ago. Since the early 1960s, when Robert Guthrie^{11,12} devised a screening test for phenylketonuria (PKU) using a newborn bloodspot dried onto a filter paper card, more than 150 million infants have been screened for a number of genetic and congenital disorders. States and territories mandate newborn screening of all infants born within their jurisdiction for certain treatable disorders that may not otherwise be detected before developmental disability or death occurs. Newborns with these disorders typically appear normal at birth. The testing and follow-up services of newborn screening programs are designed to provide early diagnosis and treatment before significant, irreversible damage occurs. Appropriate compliance with the medical management prescribed can

allow most affected newborns to develop normally. The generally acknowledged components of a newborn screening system^{4,6,13} include the following:

1. Education of professionals and parents;
2. Screening (specimen collection, submission, and testing);
3. Follow-up of abnormal and unsatisfactory test results;
4. Confirmatory testing and diagnosis;
5. Medical management and periodic outcome evaluation; and
6. System quality assurance, including program evaluation, validity of testing systems, efficiency of follow-up and intervention, and assessments of long-term benefits to individuals, families, and society.

Based on cumulative data from newborn screening programs, reported annually to the HRSA-funded NNSGRC, it is estimated that about 1 in every 800 newborns in the United States—or 5,000 of 4.1 million newborns each year—is born with a potentially severe or lethal condition for which screening and the treatment for the prevention of many or all of the complications of the condition are available. As the model for public health-based population genetic screening, newborn screening is nationally recognized as an essential program that aims to ensure the best outcome for the nation's newborn population.

NEWBORN SCREENING PROGRAMS: THE CHANGING LANDSCAPE

The infrastructure landscape.

In the United States, every State (hereafter, the term "State" will include both States and territorial jurisdictions) presently has a statute or regulation mandating or allowing public health newborn screening. As such, newborn screening is universally available in varying forms to all infants born in the United States, regardless of ability to pay or other familial factors (e.g., ethnicity, area of residence, literacy level, or language). It is important that universal access to this screening and its central public health focus are maintained, while efforts move forward to bring uniformity and equity to State screening efforts.

Since the inception of newborn screening, the conditions screened for and the systems developed for follow-up have varied among States. Due to a dearth of national newborn screening standards (aside from the National Committee for Clinical Laboratory Standards (NCCLS) "Standard on Blood Collection on Filter Paper"), guidance from the HRSA-funded Council of Regional Networks for Genetic Services (CORN) and limited advice from national advisory committees and national medical or public health professional organizations regarding newborn screening policies and conditions to be included in screening mandates, each State independently determines the conditions and screening procedures for its program.

Many States utilize advisory committees and seek input from experts and other State newborn screening laboratories

and private companies in addition to independently reviewing the available scientific evidence before making recommendations for test panels. In some States, decisions about newborn screening are in the hands of the State legislature, which controls the State public health system and its finances. Every State has a statute or regulation that allows or mandates universal newborn screening—sometimes specifying the conditions to be screened, the consent/dissent process, the laboratory, and the laboratory testing procedure to be used. In most cases, decisions about the newborn screening panel are delegated to State health officials, a State board of health, or a genetics or newborn screening advisory committee. Sometimes the decision-making process might involve a combination of agencies, advisory bodies, and policy makers.

Pilot studies usually precede the formal implementation of changes to the newborn screening panels. In addition, the mechanism to expand testing panels, change testing protocols, and fund newborn screening varies among the States, with the basic criteria from the inception of newborn screening being used by many.¹⁴ Due to these factors and a lack of national consensus or guidelines, there is presently a large disparity in screening services available to newborns. For example, at the present time, eight States mandate screening for as few as four conditions, while a number of States screen for as many as 30 conditions (information taken from NNSGRC website www.genes-r-us.uthscsa.edu/nbsdisorders.pdf July 20, 2004). This divergence among States regarding which conditions should be mandated for screening has resulted from several factors, including differences in: 1) the level of resources available (personnel, equipment and service capacity); and 2) interpretations of the available data concerning given conditions (incidence, treatability, impact) and new screening methodologies.¹⁵

Approaches to calculating the number of conditions included in screening also are variable, with some programs counting hemoglobinopathy screening as a single test and others including it as one of several tests (given the simultaneous ability to detect over 700 variant conditions including SS-disease, SC disease, $S\beta+$ -thalassemia, etc.). The expert group concluded that there should be standardization of what constitutes a screened condition. (This issue is discussed in greater detail in the section describing the conditions evaluated.)

It is clear that States must retain strong oversight of mandated screening programs in order to ensure the appropriate delivery of quality screening and ancillary services to the screened population. However, how local ancillary services are to be directly provided within programs is less clear, particularly given the nationwide lack of the specialized medical expertise and laboratory testing that is needed to definitively diagnose many of these rarer inherited genetic conditions. One suggestion to address the maldistribution of needed medical expertise has been through the organization of that expertise at the regional level, as with the newly funded HRSA/MCHB Regional Genetics and Newborn Screening Collaboratives. This effort is supported by the history of regionalization (geographically close) and consolidation (geographically dispersed) of newborn screening laboratory testing services, which has been

advantageous for States with low numbers of births. Regional programs have higher numbers of laboratory tests, which results in cost savings and decreased analytical variability.

Another challenge raised by the expansion of newborn screening is the lack of interconnecting relationships between child health professionals and subspecialists, particularly in rural areas—a problem complicated by the diversity of very rare conditions identified by the programs. There are limitations in the local availability of specific expertise for many conditions, and considerable needs exist in the areas of training and education throughout the health care system. Furthermore, improvements in the newborn screening system and the expansion of the number of conditions for which screening is offered have costs, and these costs and the associated benefits seem to accrue independently of the public and private health care delivery systems, which complicates their integration. Many States provide the programs necessary to ensure that screening and diagnosis will occur, but they are limited in their ability to ensure long-term management, including the provision of the necessary long-term treatments and services.

The societal implications of expanding newborn screening also are significant. For example, screening for additional conditions that occur with greater frequency in different ethnic groups could lead to discriminatory practices against individuals as well as the ethnic groups associated with particular disorders. In addition, difficult decisions must be made about the nature of the benefits that might be realized from newborn screening. Historically, screening has focused on conditions for which the improvement in outcome for the infant has been substantial. However, newborn screening could identify many conditions for which the improved outcomes may be more incremental, including disorders that are associated with mental retardation, such as fragile X syndrome, for which early intervention programs can improve long-term cognitive outcomes, but not with the expectation of a normal outcome.¹⁶ Finally, the nature of genetic disease is such that knowledge of its presence can be of value to other family members. Previously, this factor has not been considered by newborn screening programs.

Other considerations arise from private sector testing availability and competition. Often, private laboratories—either commercially- or university-based laboratories—offer an expanded number of conditions screened through the technologies they employ. They may provide contracted services to programs or offer additional screening for conditions not mandated in the program in the State in which the family resides. As a result, some States now mandate that all parents be informed of the availability of additional screening tests. This type of information often is delivered at the last minute and its use may not be supported by hospital staff and medical personnel. However, even though additional screening may be available when initiated by consumers, it is only through State public health that access to newborn screening for *all* babies can be assured at the present time.

The changing technological landscape

Three major technological challenges have occurred over the past few decades with regard to newborn screening. The first is the expansion of knowledge of the causes and treatment of genetic diseases. The second is the rapid expansion of diverse technologies that may be used in screening. The third is the proliferation of tiered testing strategies to enhance the positive predictive value of screening.

The sequencing of the human genome as a public/private partnership has allowed for a better understanding of the genetic bases of many diseases. This fundamental biological knowledge has led to the proliferation of new therapies stemming from intensive research efforts in both the private and public sectors. The pace of Food and Drug Administration (FDA) approval of innovative therapies has quickened. These and other factors are likely to continue to lead to an expanding panel of conditions for which newborn screening may be of benefit.

Simultaneously, there are new technological developments that allow more types of testing at reasonable cost that can be considered for application to universal newborn population screening. Examples include hearing screening, EKG screening for long QT syndrome, acylcarnitine screening, screening with molecular arrays, and screening with immunoaffinity columns. Particularly notable is the implementation of multiplex platforms that allow a single type of specimen preparation and simultaneous (or nearly simultaneous) screening for multiple different disorders. Going from one test for one disorder to one test for multiple disorders has the potential to reduce costs per condition tested and can lead to test expansion if these new technologies can be integrated safely and effectively into newborn screening programs. One potential concern associated with expansion of screening panels is the impact on follow-up testing and tracking. If the proportion of false positive cases requiring additional tests that are identified in screening laboratories rises excessively, this could undermine the acceptance of such testing by both the parental and medical communities, as well as potentially diminish the cost benefit of additional testing.

Multiplex testing technologies are emerging that can simultaneously identify multiple analytes from a single analytical process. Some multiplex testing requires that an analytical target first be identified and placed in the multiplex test (e.g., genomic arrays). Other multiplex testing provides the additional testing information without the need for specific target selection (e.g., DNA sequencing). For example, testing for hemoglobinopathies by isoelectric focusing (IEF) provides information not only about hemoglobin S, the primary target of screening, but also about more than 700 other possible hemoglobin variants, some of which may be clinically significant (e.g., Hb C and E).¹⁷

In the case of MS/MS, the multiplex testing can occur in different modes, because it is possible to operate the instrument by either selecting specific targets or analyzing full profiles.¹⁸ When used on selected targets, it is referred to as

selective reaction monitoring (SRM), which is also called multiple reaction monitoring, a process that allows for the selective evaluation of specific ion species instead of a profile within a mass range. Increasingly, MS/MS is being used in newborn screening laboratories.¹⁹ The technology is appealing for several reasons, including sensitivity for detecting ion species in low concentration, ability to quantify results relative to internal standards, high-throughput and precision, and the opportunity to simultaneously measure multiple ion species.^{15,20} However, MS/MS is a complex testing platform requiring specific training and experience in order to optimize its use.¹⁸

Although multiplex testing allows the addition of many more conditions to a screening panel, it presents a series of issues that influence the screening and health care system, ultimately affecting the screening services that might be available to the public. The availability of multiplex testing increases the number of conditions that can be considered for newborn screening that otherwise might not have been considered for screening using traditional criteria, such as incidence and treatability. Thus, our perception of screening performance characteristics is also modified. For example, multiplex technology might also reveal clinically significant conditions other than those that were the primary targets of screening but which are determined in the course of diagnostic confirmation of the screening test results. The screening laboratory may not have optimized the screening for the detection of these other conditions but they are typically part of the differential diagnosis of a primary target condition. Rather than evaluate single conditions for their inclusion in newborn screening, we must now consider how best to use the additional information revealed in the diagnostic laboratory about other related conditions.

Although information about conditions for which treatment options are scarce or not yet reported can lead to increased stresses on families and the health care system, early information can also lead to knowledge of the condition for the family, thus avoiding a potential diagnostic odyssey or inappropriate therapies. In addition, early information provides opportunity for better understanding of disease history and characteristics, and for earlier medical interventions that might be systematically studied to determine the risks and benefits. Multiplex testing and the identification of conditions falling outside of the uniform screening panel provides the opportunity for such conditions to be included in research protocols. Therefore, the criteria used to include a condition in a mandated newborn screening panel are not necessarily straightforward scientific or clinical criteria, but often involve complex ethical, legal, and social policy decisions.

Aside from new multiplex technology for screening, there has also been the introduction of tiered testing strategies to enhance the positive predictive value of screening and reduce the number of infants referred for additional testing.²¹ For example, in the United States, the primary analyte used for congenital hypothyroidism (CH) newborn screening has been thyroxin (T4), because most newborns are screened before the optimal time for screening with thyrotropin (thyroid stimulating hormone, TSH). TSH primary screening offers improved

specificity only after the period of neonatal surge and does not identify cases of central hypothyroidism. To decrease the recall rate, most screening programs have utilized a second-tier test with TSH following the identification of a certain number of increased-risk newborns through T4 initial testing.²² In such cases, secondary hypothyroidism may also be detected on the basis of the test results, even though it is not the primary target of screening. Similarly, it has been shown that the rate of false positive results in CAH screening can be significantly reduced by profiling steroids by MS/MS as a second-tier test.²³

In addition, the testing of specific DNA mutations in newborn screening (e.g., CF screening algorithms utilize a second-tier DNA mutation panel following initial screening for immunoreactive trypsinogen (IRT) and hemoglobinopathy screening algorithms that include DNA testing) can minimize the recall rates.²⁴ The testing of DNA mutations also has led to a new category that includes unaffected or minimally affected cases (e.g., carriers, benign hyperphenylalaninemia, and detection of hemoglobin Barts). Confirmation of such results and explanation of their significance can be costly. These examples highlight the ongoing process that occurs in newborn screening laboratories whereby analytes are identified that are clearly abnormal in a particular condition but still need to be analytically and clinically validated in a population screening setting.

The evidence based landscape

Assessing the evidence on conditions as to their appropriateness for newborn screening is complex, and there are limitations in the availability and interpretation of data about many of the conditions. The incidence of rare genetic diseases is often variable among different populations and can be biased by the nature of the populations involved in research and the severity of the conditions in those coming to the attention of health care professionals. Many of the conditions are ultra-rare and they may have multiple genetic etiologies. For instance, the tetrahydrobiopterin (BH4) deficiencies are a heterogeneous group of disorders that affect phenylalanine homeostasis.²⁵ BH4 deficiencies are detected as a by-product of screening for phenylketonuria due to hyperphenylalaninemia. They include disorders that affect the regeneration or biosynthesis of BH4. The condition referred to as biopterin cofactor biosynthesis defect is caused by one of two genes—GTP cyclohydrolase I (GTPCH) and 6-pyruvoyl-tetrahydrobiopterin synthase (PTPS)—and the condition referred to as biopterin cofactor regeneration defect is caused by one of two genes—pterin-4 α -carbinolamine dehydratase (PCD) and dihydropteridine reductase (DHPR). Due to the biochemical similarities of the deficiencies resulting from blocks in these interrelated pathways, the clinical courses are similar in those with the typical severe forms of GTPCH, PTPS, and DHPR deficiencies. Approximately 57% of the rare BH4 abnormalities involve PTPS deficiency. However, due to the similarities in phenotype and treatment, the BH4 abnormalities are commonly combined with the two aforementioned groups and the treatments are similar. Hence, incidence as it relates to the genetic etiology is usually combined for the

two subtypes. Treatment for the conditions is related to the degree of hyperphenylalaninemia and to the degree of impairment of biogenic amine production, which varies among those affected. Further, a treatment involving BH4 administration is now approved in Europe, following clinical trials, that demonstrated that both GTPCH and PTPS are responsive to BH4. Due to the fact that GTPCH is very rare, yet quite similar to PTPS, the affected are aggregated when treatment is assessed. In any case, due to the rarity of these conditions, it is not until a very large general population has been identified through screening that penetrance and expressivity of disease are determined and a true incidence figure becomes available. In order to ensure that new therapies for these rare and severe genetic diseases will be available, regulatory agencies sometimes accept premarket evidence from smaller treatment groups while shifting the burden for the collection of additional information to FDA Phase IV postmarket surveillance, as was reported in FDA News for Fabryzyme[®] for the treatment of Fabry disease. (See <http://www.fda.gov/bbs/topics/NEWS/2003/NEW00897.html>)

Having such treatments available earlier means that it becomes increasingly difficult to collect information on the natural history of the untreated condition. In fact, there has not been a natural history study of PKU conducted since the 1970s because the affected infants are routinely identified in screening are treated, respond well to the treatment. Understanding the genetic basis of these conditions has led to this relatively rapid transition between ability to diagnose and the development of treatments based on the underlying biology and pathology of genetic diseases, particularly those that involve the replacement of defective enzymes. Hence, it becomes increasingly important to develop national systems for the collection of clinical information about those individuals identified in screening to further inform our understanding of the screened conditions and to further evaluate treatment modalities through an iterative process.

The assessment of the evidence on the performance characteristics (analytical and clinical sensitivity, specificity, and positive predictive values) of the tests, as used in newborn screening is complex. Many of the screening tests use technologies that are the gold standard in the diagnostic setting, such as HPLC or IEF for hemoglobinopathies or MS/MS for the acylcarnitine disorders. Although one can demonstrate very strong analytical and clinical performance in a diagnostic setting, clinical performance in screening is a function of the cut-offs that are used by the screening laboratories to capture the most affected persons. States often assign varying cut-offs to analyte levels and often use different screening test algorithms, including second-tier tests or repeat tests to arrive at a determination of whether the specimen is within the normal range, with highly variable case definitions at screening. This lack of standardization makes it quite complex to assign a level of performance to the screening tests at a national level or to compare the performance of programs.

Finally, the evidence base for newborn screening is complicated by the differing views of the interest groups involved. For purely scientific and medical issues, the scientific literature

provides objective information about different aspects of conditions, such as incidence, treatment efficacy, and diagnostic confirmation. However, some criteria have significant subjective aspects that require the consideration of more than just scientific and expert opinion. Cost is an example of a subjective criterion because it is a contextual concern and can only be measured against the value of the outcome. Other criteria may be perceived differently by the professional community or by other nonscientific or nonmedical interest groups. For example, parents often consider difficult the impact of treatments that health care professionals consider to be simple (e.g., maintaining a child on a specified diet). Some criteria are perceived differently among varying groups of professionals. For example, primary health care professionals in urban areas often have greater access to subspecialists than do those in rural areas. It is often difficult to balance the scientific evidence against the values that different groups place on newborn screening to reduce mortality and morbidity of diseases.

The need for evaluation of newborn screening systems

The lack of equitable newborn screening services offered for infants, the changing dynamics of emerging technology, and the complexity of genetics require an assessment of the state of the art in newborn screening and a perspective on the future directions such programs could take. In addition, programs must include an assessment of the availability of needed resources, both public and private, when determining which conditions should be included. A national, organized approach to differentiating among these many competing needs would help create a more informed process for deciding what tests should be included in newborn screening programs.

Since the first State newborn screening programs began, periodic assessments have been made. As early as 1968, the World Health Organization (WHO) issued a report urging that screening tests be appropriate and straightforward.²⁶ In 1975, the National Academy of Sciences (NAS) redefined genetic screening and established the fundamental principles and rules of procedure for genetic testing (these did not vary significantly from the 1968 WHO recommendations). NAS also made recommendations regarding the aims of testing and screening, criteria for testing, and the quality of testing.¹³ In 1997, the Task Force on Genetic Testing, created by the National Institutes of Health-Department of Energy Working Group on Ethical, Legal and Social Implications of Human Genome Research, focused on the quality of testing and recommended that screening tests demonstrate analytical and clinical validity and utility²⁷ (Holtzman and Watson, 1997 available at <http://www.genome.gov/10001733>). In 1999, at the request of HRSA, AAP convened a Newborn Screening Task Force that provided a comprehensive evaluation of the current state of newborn screening programs in the United States.¹³ The Task Force recommendations covered the public health and clinical care system, the roles of professionals and the public, issues of disease surveillance and research, and the economics of newborn screening. The report recommended that “HRSA should engage in a national process involving gov-

ernment, professionals, and consumers to advance the recommendations of this Task Force and assist in the development and implementation of nationally recognized newborn screening system standards and policies.” In addition, the AAP Task Force¹³ thought that greater uniformity would benefit families, health care professionals, and the newborn screening programs. In 2000, the March of Dimes, an organization that has advocated on behalf of newborn screening programs, recommended that tests be rapid, high quality, and accurate and that cost should not be a major consideration.²⁸ Subsequently, the March of Dimes recommended that all States screen for nine conditions plus newborn hearing loss (see www.marchofdimes.com/professionals/580.asp).

B. Methods used for assessing conditions

As an initial step in the process, ACMG convened a newborn screening expert group that included participants with expertise in various areas of subspecialty medicine, primary care, health policy, law, ethics and public health, and consumers. The expert group also formed two expert work groups to provide more in-depth analysis in two specific areas: the uniform panel and its criteria, and the diagnosis and follow-up system. Members of the expert group and work groups are listed at the beginning of this report. Work group members were selected based on their abilities to bring a strong scientific and clinical—rather than organizational—perspective to the issues under consideration. Not only were efforts made to ensure cultural, ethnic, and geographic diversity, there also were efforts to involve health care professionals and other interested parties from a wide range of fields and backgrounds, including expert representation from public health laboratories and program administration; individuals who are involved in the delivery of specialty care; primary care and nonphysician health care professional groups that are involved with the patients and families; and parents who have been directly affected by newborn screening programs.

The project depended on a variety of types of input obtained through expert reviews of the scientific literature, presentations from international and national invitees at six meetings of the expert group, solicitations for public and professional comment, and detailed assessments provided by the work groups. Considerable information was acquired through the use of disease-specific surveys that were broadly distributed and augmented by direct requests for input from acknowledged experts for the conditions under consideration. Areas in which deficiencies were found in the information available in the scientific literature were identified as well.

The expert group followed a two-tiered approach to assessing conditions that allowed for the views of experts of various types, including consumers, to be considered while still deferring to the evidence in the scientific literature. In the first level of the assessment, the expert group sought broad input through a survey of individuals and organizations with an interest in newborn screening. The expert group utilized a data collection instrument, distributed through a survey and directly to experts, to seek unpublished data and views related to

the criteria by which conditions were to be evaluated. The opinions of experts and others were quantified using the scoring system assigned to each criterion in the data collection instrument. Conditions were then placed preliminarily into categories reflecting their overall scores on the data collection forms. In the second level of the assessment, the scientific and medical evidence bases relating to the conditions under consideration were developed. Each condition was then reassessed to ensure that the evidence base confirmed that three critical evaluation categories were met in order to define a uniform panel of conditions to be targeted by newborn screening programs.

Establishing principles for the development of newborn screening guidelines

Many factors could influence a decision to include a given condition in a newborn screening program, including, for example, the severity of the condition, the availability of effective treatment, the age of onset, and the complexity or cost of the test.²⁹ In developing the criteria to evaluate conditions and make recommendations, the expert group relied on a set of basic principles developed at the onset of the project. The order of these principles is not intended to suggest a prioritization.

An overarching concept is utility—that is, an approach that delivers the greatest good to the greatest number of people, while recognizing the need for some flexibility and the use of alternative mechanisms by screening programs. Newborn screening policies and practices have national, regional, and local implications. Although national uniformity is a goal for newborn screening programs, there also may be a need, in limited and specific circumstances (such as meeting local and community public health needs), to screen for certain genetic conditions identified only in given populations.

Newborn screening involves many parties. In addition to the child and his or her family or guardian, these include public health officials, health care professionals, private insurers, government officials, researchers, policymakers, educators, and others. This report seeks to acknowledge the full range of participants involved.

1. Universal newborn screening is an essential public health responsibility that is critical to improve the health outcome of affected children.

To ensure that all United States newborns have access to screening and to promote a systems approach to population-based health care, it is critical that newborn screening remain a public health function.

2. Newborn screening policy development should be primarily driven by what is in the best interest of the affected newborn, with secondary consideration given to the interests of unaffected newborns, families, health professionals, and the public.

A key factor determining the inclusion of particular conditions in newborn screening programs is the potential for the affected newborn to realize a significant improvement in quality of life as a result of the screen-

ing. Although the expert group gives primary consideration to newborns that are being screened, it is clear that many others are also affected by newborn screening. Newborns that do not screen positive can benefit from the elimination of certain diagnoses, and families benefit independent of the newborn that was screened. Furthermore, because these programs can decrease mortality and morbidity, public health professionals, the public, and the health care system may derive benefits from newborn screening programs, such as cost reductions for overall health care services. There may also be negative consequences for newborns and families that result from screening, including the potential negative impact of a false-positive screening result. Aside from the financial cost of a medical work-up to confirm that a suspected condition does not exist, there may be associated anxiety and stress for the family.

3. Newborn screening is more than testing. It is a coordinated and comprehensive system consisting of education, screening, follow-up, diagnosis, treatment and management, and program evaluation.

To realize the benefits from newborn screening, all components of the program must function well together. The six critical components of newborn screening programs—education, screening, follow-up, diagnosis, treatment and management, and evaluation—are important to the overall functioning of individual newborn screening programs and the system in which they operate.³⁰ There must be assurance of timely and accurate reporting and tracking of abnormal results. In order to know whether a program is functioning effectively and efficiently, it is important to know whether the expected health benefits are being realized.

4. The medical home and the public and private components of screening programs should be in close communication to ensure confirmation of test results and the appropriate follow-up and care of identified newborns.

The medical home concept has evolved as the central focus for the care of patients in their communities and should be the center of communication, primary care, and coordination of care for individuals.³¹ There is increased recognition that enhanced communication between the clinical care system and public health programs is necessary to ensure optimal care and outcomes for the affected newborns. It is essential to establish close communication among State public health programs, the newborn's medical home, and the subspecialists commonly involved in the diagnosis and follow-up of affected newborns.

5. Recommendations about the appropriateness of conditions for newborn screening should be based on the evaluation of scientific evidence and expert opinion.

There are ever-increasing numbers of relatively rare conditions for which clinical knowledge is rapidly growing but for which the published literature may be

sparse or outdated. Moreover, clinical expertise in treating many of these conditions may be limited. Given that all screening programs must rely on the same published knowledge base and a limited number of experts, a national process of scientific evaluation seems most practical. As new evidence emerges and opinions change, there should be a system in place for prompt review and release of updated recommendations.

In 2003, the Secretary's Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children was established by the Department of Health and Human Services (DHHS). Its mandate was to advise and guide the Secretary of DHHS regarding the most appropriate application of universal newborn screening tests, technologies, policies, guidelines, and programs in order to effectively reduce morbidity and mortality in newborns and children who have or who are at risk for heritable disorders. The committee's purpose is to provide the Secretary with: "advice and recommendations concerning the grants and projects and technical information needed to develop policies and priorities that will enhance the ability of State and local health agencies to provide for newborn and child screening and counseling and health care services for newborns and children having or at risk for heritable disorders." (Available at <http://mchb.hrsa.gov/programs/genetics/committee/>)

6. To be included as a primary target condition in a newborn screening program, a condition should meet the following minimum criteria:

- It can be identified at a period of time (24 to 48 hours after birth) at which it would not ordinarily be clinically detected.
- A test with appropriate sensitivity and specificity is available.
- There are demonstrated benefits of early detection, timely intervention, and efficacious treatment.

Determining the appropriateness of a condition for newborn screening is a complex process. Although the emergence of new technologies such as MS/MS has altered views of which conditions should be included in mandated screening programs, in this report the primary targets of screening are those that meet the three criteria previously specified. A secondary target is one that is identified while searching for the primary target (e.g., HbC results from IEF while looking for HbS) or a clinically significant condition that is likely to be detected when performing a comprehensive profile of a given group of biochemical markers (e.g., GA2 may be identified while determining MCAD status (C8 acylcarnitine is elevated in both)).

7. The primary targets of newborn screening should be conditions that meet the criteria listed in #6 above. The newborn screening program should report any other results of clinical significance.

Many technologies can be applied to screening for pri-

mary targeted conditions. Some allow for more than one condition to be identified in a single procedure, and some provide important information about the presence of conditions that may not meet all of the criteria needed to be considered a primary target condition. The advent of molecular arrays and MS/MS has significantly broadened this potential. It is not necessarily the responsibility of the screening program to monitor the long-term follow-up of patients identified with clinically significant conditions that are not the primary targets of newborn screening. However, the significant costs of the diagnostic odysseys that may ensue following the birth of a child whose condition may have been suspected based on newborn screening results, and the related costs to families and the system of introducing futile therapies might be avoided if clinically significant results from newborn screening programs are shared with the newborn's primary caretaker.

8. Centralized health information data collection is needed for longitudinal assessment of disease-specific screening programs.

Mechanisms and systems that allow for the collection of short- and long-term data on affected individuals while protecting their right to privacy will allow for assessment and improvement of program performance and individual health outcomes. The pooling of information about health outcomes, treatment protocols, case definitions, and diagnosis and confirmation algorithms will improve care for the infants identified in the programs. Furthermore, it is often difficult to ascertain the natural history of rare diseases because of their low frequency and because they often exhibit genetic variability in severity and expression. Hence, data collection and shared data evaluation can significantly inform program decision-making and medical science. General population data are also needed to better understand certain approaches to screening (e.g., genomics), where the variability in expression of mutations is not entirely clear until individuals without the classical presentation of a condition are tested.

9. Total quality management should be applied to newborn screening programs.

As with any programmatic effort, improvements result from careful and continuous monitoring of key steps in the process, the assessment of that information, and the introduction of changes that continuously improve program performance. Uniform and consistent monitoring of system quality indicators can provide information about the relative performance of screening programs.

10. Newborn screening specimens are valuable health resources. Every program should have policies in place to ensure confidential storage and appropriate use of specimens.

Specimens obtained for newborn screening have tre-

mendous long-term value. They can be used for purposes of program quality management, to help inform deliberations about program expansion, for research on testing technology and treatment, and for epidemiologic studies. This is not to imply that every State should store all specimens forever but, rather, that there should be a sufficient number of States with diverse populations and long-term storage of residual specimens to provide this critical resource. Regardless, it is important to ensure the confidentiality of those persons whose specimens are stored. The use of specimens for nontherapeutic purposes must not alter the willingness of the public to participate in newborn screening programs and related activities.

11. Public awareness, coupled with professional training and family education are significant program responsibilities that must be part of the complete newborn screening system.

Because newborn screening can have a significant impact on health outcomes for affected newborns, it is essential that the public as well as health care and public health professionals be informed of the availability of the programs and of changes that are made. Education and awareness are essential to both the quality of the screening programs and participation by the public and by health care professionals. As such, information sharing and education are critical program responsibilities.

Choosing the conditions

Eighty-four conditions were evaluated using these criteria (see Table 1). The conditions were chosen for several reasons. Any condition currently included in private, State, or national newborn screening programs was considered. Other conditions were included because they are known to be coincidentally revealed by some of the technologies used in newborn screening. Still others were identified by members of the public, the expert group, and work groups as worthy of consideration because they are important from a public health standpoint and/or there is a high level of public and/or scientific interest in screening for the condition. Hemoglobinopathy screening was mainly driven by the conditions associated with a hemoglobin S allele. Among these, Hb SS, Hb SC and Hb S β -thalassemia were considered separately. Variant hemoglobinopathies included other conditions associated with an Hb S allele. Additional hemoglobinopathies revealed by screening, such as Hb E, are not the conditions to which screening currently is targeted. As discussed below, compromises were made in the lumping or splitting apart of conditions to be listed for assessment.

To a limited extent, the conditions listed as considered by the expert group represent a compromise among the various options. The intent was to distinguish many of the more common forms of the condition from others though there are still situations in which some very rare conditions are subsumed under a more general name for the condition.

The group considered it important to fully assess all conditions and to ensure that any apparent deficiencies were properly recognized so that disease-specific advocacy groups and the research community could focus on these deficiencies in developing their research agendas.

Developing evaluation criteria and their comparative values

Generally, a medical condition is assessed by itself to determine whether it should be included in a public health newborn screening program,^{14,29} rather than being assessed along with a number of other conditions in a way that would allow for comparative ranking. Historically, this is primarily because individual conditions have been identified by individual testing platforms. Although conditions have usually been compared on the basis of relative incidence, there was little need for additional discriminating criteria given the general availability of traditional testing methodologies and treatments. Thus, comparative analyses of screened conditions or evaluations of the scientific evidence for or against inclusion of the conditions have not been formally conducted nationally, though this has often been done within individual programs.

Until recently, the capability of the currently available testing technology limited the conditions that could reasonably be included in a screening panel. Now, however, new information emerging from the clinical and scientific literature, combined with evolving technologies, has made it possible for increasing numbers of rare conditions to be detected simultaneously from single screening tests, making it reasonable to attempt more complex relative comparisons when deciding on conditions that should be added to a screening panel. Thus, it is no longer a simple matter to decide which condition should be added to a screening panel based on incidence, when a group of conditions may be simultaneously detected from a single analytical procedure and the group incidence (or impact to society) may be of higher relative importance than any of the single conditions within the group. In addition, even if multiple conditions could be detected, the question of whether they *should* be detected remains, when, for example, no efficacious treatment exists. Increasing the complexity of this decision-making process is the fact that all of the conditions detected may not have similar clinical outcomes for all children.

In recent years, professional groups in other countries have attempted to develop an organized, national approach to determining which conditions should be included in newborn screening panels. The Health Technology Assessment Program of the National Health Service of the United Kingdom has initiated a national program to systematically review the scientific and medical literature on inborn errors of metabolism, neonatal screening technology, and screening programs. Their goal is to analyze the costs and benefits of introducing MS/MS-based screening of amino acid disorders, fatty acid oxidation defects, and organic acid disorders, as well as other conditions screened on an individual test basis within the United Kingdom health system.¹⁰ This extensive analysis assigned weights to various aspects of specific conditions and their associated

Table 1
Individual conditions considered in the data collection instrument

Group	Condition	Code	
Endocrinology	Congenital adrenal hyperplasia	CAH	
	Congenital hypothyroidism	CH	
	Diabetes mellitus, insulin dependent	IDDM	
Hematology, Hemoglobinopathies	Hb SS disease (Sickle cell anemia)	Hb SS	
	Hb S/C disease	Hb S/C	
	Hb S/ β -thalassemia	Hb S/ β -Thal	
	Other variant Hb-pathies (including Hb E)	Var Hb	
	Glucose-6-phosphate dehydrogenase deficiency	G6PD	
Infectious Diseases	Human HIV infection	HIV	
	Congenital toxoplasmosis	TOXO	
	Congenital cytomegalovirus infection	CMV	
	Alpha 1-antitrypsin deficiency	A1AT	
	Adenosine deaminase deficiency	ADA	
	Biliary atresia	BIL	
	Cystic fibrosis	CF	
	Duchenne and Becker muscular dystrophy	DMD	
	Familial hypercholesterolemia (heterozygote)	FHC	
	Miscellaneous Genetic Conditions	Fragile X	FX
Hearing loss		HEAR	
Hyperbilirubinemia*		HPRBIL	
Neuroblastoma		NB	
Severe combined immunodeficiency		SCID	
Turner syndrome		TURNER	
Wilson disease		WD	
Amino Acid Disorders		Phenylketonuria	PKU
		Benign hyperphenylalaninemia	H-PHE
		Defects of bipterin cofactor biosynthesis	BIOPT BS
		Defects of bipterin cofactor regeneration	BIOPT REG
		Homocystinuria	HCY
		Hypermethioninemia	MET
	Maple syrup (urine) disease	MSUD	
	Tyrosinemia type I	TYR I	
	Tyrosinemia type II	TYR II	
	Tyrosinemia type III	TYR III	
	Carbamylphosphate synthetase deficiency	CPS	
	Ornithine transcarbamylase deficiency	OTC	
	Citrullinemia	CIT	
Citrullinemia type II	CIT II		

(continued)

Inborn Errors of Metabolism

Table 1
Continued

Group	Condition	Code		
Inborn Errors of Metabolism	Argininosuccinic acidemia	ASA		
	Argininemia	ARG		
	Classic galactosemia	GALT		
	Carbohydrate Disorders	Galactokinase deficiency	GALK	
		Galactose epimerase deficiency	GALE	
		Congenital disorder of glycosylation type Ib	CDG Ib	
		Fatty Acid Oxidation Disorders	Carnitine uptake defect	CUD
			Carnitine palmitoyltransferase Ia deficiency (L)	CPT IA
	Carnitine palmitoyltransferase Ib deficiency (M)		CPT IB	
	Carnitine/acylcarnitine translocase deficiency		CACT	
	Carnitine palmitoyltransferase II deficiency		CPTII	
	Very long-chain acyl-CoA dehydrogenase def.		VLCAD	
	Long-chain 3-OH acyl-CoA dehydrogenase def.		LCHAD	
	Trifunctional protein deficiency		TFP	
	Dienoyl-CoA reductase deficiency		DE-RED	
	Glutaric acidemia type II		GA2	
	Medium-chain acyl-CoA dehydrogenase deficiency		MCAD	
	Medium/short-chain 3-OH acyl-CoA DH def.		M/SCHAD	
	Medium chain ketoacyl-CoA thiolase deficiency		MCKAT	
	Short-chain acyl-CoA dehydrogenase deficiency	SCAD		
	Lysosomal Storage Diseases	Fabry disease	FABRY	
		Krabbe disease	KRABBE	
		Pompe disease	POMPE	
	Hurler-Scheie disease	MPS-1H		
	Organic Acid Disorders	Lysosomal storage diseases	LSD	
		Propionic acidemia	PA	
		Multiple carboxylase deficiency (Holocarboxylase Synthetase deficiency)	MCD	
		Methylmalonic acidemia (mutase)	MUT	
		Methylmalonic acidemia (Cbl A, B)	Cbl A,B	
		Methylmalonic acidemia (Cbl C,D)	Cbl C,D	
		Isobutyryl-CoA dehydrogenase deficiency	IBG	
		2-Methylbutyryl-CoA dehydrogenase deficiency	2MBG	
		2-Methyl 3-hydroxy butyric aciduria	2M3HBA	
		β -Ketothiolase deficiency	β KT	
		Isovaleric acidemia	IVA	
		3-Methylcrotonyl-CoA carboxylase deficiency	3MCC	
3-Methylglutaconic aciduria		3MGA		
3-hydroxy 3-methyl glutaric aciduria		HMG		
Glutaric acidemia type I		GA I		
Malonic aciduria	MAL			

(continued)

Table 1
Continued

Group	Condition	Code
Other IEM	Biotinidase deficiency	BIOT
	X-linked Adrenoleukodystrophy	ALD
	Smith-Lemli-Opitz syndrome	SLO
	Guanidinoacetate methyltransferase deficiency	GAMT
	Arginine: glycine amidinotransferase deficiency	AGAT
	Creatine transporter defect	CR TRANS

NOTE: Neonatal hyperbilirubinemia (Kernicterus) (code HPRBIL) was added to this list after the completion of the data collection instrument.

tests and treatments, and assigned a qualitative value to the published information available. This effort has highlighted the difficulties inherent in attempts to balance costs and benefits against the value that the public and families place on such screening.

The Human Genetics Society of Australasia developed criteria for placing conditions into one of four tiers. These tiers are determined by the nature of the benefit of the screening to the newborn, the benefit of the screening balanced against the cost, the suitability of the test, and the availability of appropriate and organized diagnostic and follow-up services (available at <http://www.hgsa.com.au/Word/HGSApolicyStatementNewborn-Screening0204-18.03.04.doc>).

More recently, Belgium has sought to assign values to the Wilson and Jungner criteria,¹⁴ in order to weigh conditions against each other (see Box 1). Although novel, this system was considered to be less detailed than needed because many of the Wilson and Jungner criteria are subjective and therefore less amenable to the application of a metric and therefore quantification.

In the United States, several states, including Nebraska, Tennessee, and Washington, recently developed criteria and systems for assessing and comparing conditions. With the establishment of the 2003 federal Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children, the potential for development of national policies and recommendations should lead to a more uniform or equitable approach to newborn screening.

None of the existing systems allowed for adequate comparative analysis of conditions being considered for newborn screening. Further, the evolution of screening programs and the screening technologies used have added new variables to be considered when assessing conditions. The ACMG expert group chose to develop a modified system for the assessment of conditions for their appropriateness for newborn screening.

The Uniform Panel Work Group developed the data collection instrument to use during the project's first phase to quantitatively evaluate the features of conditions under consideration for inclusion in a potential uniform screening panel. Using a weighted scoring system, the conditions were evaluated according to criteria in three main categories:

1. The clinical characteristics of the condition;
2. The analytical characteristics of the test; and

Box 1 Wilson and Jungner Criteria for Appraising the Validity of a Screening Program

1. The condition being screened for should be an important health problem.
2. The natural history of the condition should be well understood.
3. There should be a detectable early stage.
4. Treatment at an early stage should be of more benefit than at a later stage.
5. A suitable test should be devised for the early stage.
6. The test should be acceptable.
7. Intervals for repeating the test should be determined.
8. Adequate health service provision should be made for the extra clinical workload resulting from screening.
9. The risks, both physical and psychological, should be less than the benefits.
10. The costs should be balanced against the benefits. SOURCE Wilson, J.M., and G. Jungner. *Principles and Practice of Screening for Disease*. (Public Health Paper Number 34.) Geneva: World Health Organization, 1968.

3. Diagnosis, follow-up, treatment, and management of the condition.

Within each of these categories, 19 component criteria including six characteristics of the analytical tests were considered for assigning a comparative value, or score. Conditions already included in newborn screening programs were used to model the scoring system. Each of the criteria was weighted to reflect the presumed importance of the particular criteria to the overall assessments of conditions. Experts in the conditions under consideration for newborn screening were then asked to consider the criteria and the extent to which they cover the range of issues that arise among disparate types of conditions. They were also asked to

consider whether appropriate weights were assigned to criteria, thereby acknowledging the criteria considered most relevant. The language describing the criteria and the scores associated with the range of responses to the criteria were adjusted by the expert group (see Table 2 for the criteria and the possible scores). Then, the weight accorded to each criterion was revised (i.e., the highest possible score within each category was the same). The criteria that were identified within each category were assigned a range of possible responses and related scores ranging from 0 to a maximum score that varied according to each criterion's overall importance. Conditions already included in newborn screening programs were then assessed for their performance in the system. Results were compared with those obtained by other systems de-

veloped for this purpose to determine whether the outcomes were similar.

The scoring system recognizes the strengths and limitations found in each condition and summarizes them in a ranking system. Thus, a low score in a particular area does not necessarily mean that screening for that condition will never be conducted. In fact, low scores could be radically overruled by scientific evidence of new advances in testing and treatment and should be recognized as opportunities for targeted clinical or basic research endeavors and subsequent reconsideration of the condition for inclusion.

The criteria that were developed to differentiate the appropriateness of conditions for newborn screening include some

Table 2
Combined criteria and distribution of scores in the data collection instrument (Highest possible score: 2100)
I. Condition/Disorder (subtotal score 700)

Criterion	Categories in criterion	Score
Incidence of condition	>1:5x000	100
	>1:25,000	75
	>1:50,000	50
	>1:75,000	25
	<1:100,000	0
Signs and symptoms clinically identifiable in the first 48 hours	Never	100
	<25% of cases	75
	<50% of cases	50
	<75% of cases	25
	Always	0
Burden of disease (natural history if untreated)	Profound	100
	Severe	75
	Moderate	50
	Mild	25
	Minimal	0
Individual benefits of early intervention	Clear scientific evidence that early intervention resulting from screening optimizes outcome	200
	Some scientific evidence that early intervention resulting from screening optimizes outcome	100
	No scientific evidence that early intervention resulting from screening optimizes outcome	0
Familial and societal benefits of early intervention	Early identification provides clear benefits to family and society (education, understanding prevalence and natural history, cost effectiveness)	100
	Early identification provides some benefits to family and society	50
	No evidence of benefits	0
Early diagnosis and treatment prevent mortality	Yes	100
	No	0

Table 2
Continued
II. Screening Test (subtotal score 700)

Criterion	Categories in criterion	Score
Does a sensitive AND specific screening test algorithm currently exist?	Yes	200
	No	0
Test characteristics (Yes = apply score; No = 0)	Doable in neonatal bloodspots OR by a simple, in-nursery physical method	100
	High throughput (>200/day/FTE)	50
	Overall analytical cost <1\$ per test per condition	50
	Multiple analytes relevant to one condition are detected in same run	50
	Other conditions identified by same analytes	50
	Multiple conditions detected by same test (multiplex platform)	200

III. Treatment & Management (subtotal score 700)

Criterion	Categories in criterion	Score
Availability of treatment (*)	Treatment exists and is widely available in most communities	50
	Treatment exists but availability is limited	25
	No treatment available or necessary	0
Cost of treatment (*)	Inexpensive	50
	Expensive (>\$50,000/patient/year)	0
Potential efficacy of existing treatment	To prevent ALL negative consequences	200
	To prevent MOST negative consequences	100
	To prevent SOME negative consequences	50
	Treatment efficacy not proven	0
Diagnostic confirmation	Providers of diagnostic confirmation are widely available	100
	Limited availability of providers of diagnostic confirmation	50
	Diagnostic confirmation is available only in a few centers	0
Acute management	Providers of acute management are widely available	100
	Limited availability of providers of acute management	50
	Acute management is available only in a few centers	0
Simplicity of therapy	Management at the primary care or family level	200
	Requires periodic involvement of a specialist	100
	Requires regular involvement of a specialist	0

NOTE: The two criteria marked with (*) above were combined in the data collection instrument, a score of 100 was attributed to a treatment that is inexpensive and widely available, 50 if expensive or limited availability, 0 if expensive and limited availability. The final version was prompted by feedback from several survey respondents who felt that not all options were actually considered (e.g., no treatment necessary).

that have a highly objective scientific basis and others that are more subjective. To the extent possible, the expert group relied on the scientific literature to provide the information on which its recommendations are based. Survey respondents were provided with the data collection instrument, questionnaires about the criteria themselves, the weight assigned to criteria, and the distribution of scores within a criterion. The respondents were asked to provide information on both objective and

subjective criteria as a way of determining a respondent's familiarity with the condition(s).

THE THREE MAIN CATEGORIES AND THEIR CRITERIA

Clinical characteristics of the condition

Three criteria were developed for this category:

1. Incidence Of The Condition

The incidence of the various conditions varies widely. In terms of public health importance, the more common the condition, the higher the justification for screening. Accordingly, any condition with a documented (or estimated) incidence of 1:100,000 or less received a score of zero, while an incidence of 1:5,000 or more received a score of 100. When technology allows for the condition to be detected in the course of screening for other conditions, points were added back through the appropriate testing criteria. (See “Screening Test: Availability and Characteristics,” below.)

2. Clinically Identifiable Signs And Symptoms In The First 48 Hours

In the context of public health, it is more important to screen for conditions that generally would not be detected in the newborn period based solely on routine clinical evaluation. However, it is important to recognize that there could be differences of opinion regarding whether a particular phenotype could be recognized by a typical health care provider and/or specialist, and the phenotypic variability expected among newborns with a particular condition must be considered. Nonetheless, if clinical symptoms are never detectable within 48 hours after birth, the condition received a score of 100. If clinical manifestations are always detectable, the condition received a score of zero.

3. Burden Of Disease (Natural History If Not Treated)

This is an important criterion for prioritizing the use of public health resources because it favors screening for conditions that constitute greater burdens to those affected (if the burden is profound, for example, a score of 100 was given). It is recognized that some conditions have a wide range of severity and that the test may not necessarily discriminate the milder forms from the more severe forms.

The screening test: availability and characteristics

Seven criteria are included in this category:

1. Availability Of A Sensitive And Specific Test Algorithm

This criterion is a central consideration when assigning a test or a condition to a uniform screening panel. The expert group chose to define this criterion as a test algorithm because some tests might require that additional analytes or second-tier tests be incorporated to achieve sufficient specificity (e.g., the use of T4 and TSH for the screening of CH or the use of a second-tier molecular test to improve the specificity of the IRT test for CF). This criterion was considered the first step in a decision tree without which further consideration for inclusion in newborn screening would not be possible. One hundred points were allotted to this feature of a condition. If a condition had no sensitive and specific test available that could be used in population screening, it was assigned a score of zero. However, it is acknowledged that there is

no agreed-upon level of sensitivity and specificity and that this may vary with the burden of the condition and its importance for screening.

2. Ability To Test On Either Neonatal Bloodspots Or An Alternative Specimen Type Or By A Simple, In-Nursery Procedure

Value was assigned if a test can be done on a dried blood-spot, which is a highly stable specimen type already integrated into newborn screening and on which many tests can be performed. Equal consideration was given to a screening test that could be conducted using a simple procedure or method, as with hearing screening, that would be appropriate for population screening. One hundred points were allotted to this feature of a test.

3. Test Is Based On A Platform That Offers High-Throughput Capability

Value was placed on the ability of a technology to operate in a high-throughput format that allows testing of at least 200 specimens per full-time employee equivalent per day. The ability to test a large number of specimens in a short time offers cost savings to programs and increases efficiency, both important for public health screening. Fifty points were allotted to this criterion.

4. Cost Of Test Is Less Than \$1 Per Infant Screened

Value was placed on low-cost technologies. Cost was based on the personnel, reagents and other costs associated with testing only. Differences in the scoring of conditions detected by MS/MS were likely due to higher costs when a multiplex technology is used to screen for only a few conditions rather than for a larger number of conditions. Fifty points were allotted to this feature of a test.

5. Multiple Analytes Relevant To One Condition Can Be Detected In The Same Run

The ability to detect multiple markers of a given condition within the same test increases the specificity of the method by allowing the calculation of ratios that have been shown to improve the differentiation between true positives and potential false positives. Fifty points were allotted to this feature of a test.

6. Other Conditions (Secondary Targets) Can Be Identified By The Same Analytes

Value was assigned to the ability of a test to provide information about multiple conditions using the same analyte(s). Although these secondary targets may not independently meet all of the other criteria for inclusion in the uniform screening panel, they add value to the primary target condition because their detection constitutes a clinically significant result leading to tangible benefits to the affected newborn, family, and society. Fifty points were allotted to this feature of a test.

7. Multiple Conditions Can Be Detected By The Same Test (Multiplex Platform)

Technology can add value to testing, particularly if it provides the ability to screen for many conditions in a single test. This can have public health importance above and beyond the features of the disease itself (i.e., by detecting

secondary conditions). This capability resides in technologies such as MS/MS, IEF, and HPLC for hemoglobin variants, DNA arrays used in sequencing, and labeled bead technologies. Technologies with multiplexing capability offer improved efficiency and cost-effectiveness to programs. Because of the public health importance of technologies with multiplex capabilities, this criterion was allotted two hundred points.

Diagnosis, follow-up, treatment, and management

Nine criteria were developed to assess the combined aspects of diagnostic confirmation and treatment and management:

1. Availability Of Treatment

The availability of treatment is considered an important criterion for conditions in a core newborn screening panel. Fifty points were allotted to this feature of a condition, though additional value is assigned later depending on the effectiveness of the treatment.

2. Cost Of Treatment

The cost of treatment is an important consideration in newborn screening. Although this criterion does not necessarily differentiate cost from value, it should be factored into decision-making. Fifty points were allotted to this feature of the treatment.

3. Potential Efficacy Of Existing Treatment

More effective preventive or therapeutic interventions for a given condition increase the value of testing. For many conditions, treatments could result in near normal or normal outcomes. For others, the treatment may affect only a subset of the negative phenotypes possible or allow for only incremental improvements in optimal outcome. Moreover, treatment might not be equally effective in all individuals. This was considered a critical criterion and was assigned a value of 200 points.

4. Individual Benefits Of Early Intervention

This criterion is important because the benefit to the child being screened is the overriding consideration. This was considered an objective criterion based on the quality of available evidence showing that early intervention optimizes outcome. Two hundred points were allotted to this feature of a treatment.

5. Familial And Societal Benefits Of Early Identification

Early identification of an infant with a condition can bring benefits to families and/or society beyond the prospect of treatment. Because so many of the conditions detected through newborn screening are genetic, families can benefit from establishing that there may be a genetic risk to others in the family. Society could benefit by a reduction in medical diagnostic odysseys that are costly to the health care system. One hundred points were allotted to this feature of a condition.

6. Prevention Of Mortality Through Early Diagnosis And Treatment

Prevention of mortality was assigned a value indepen-

dent of reduction of morbidity. One hundred points were allotted to this feature of a condition.

7. Availability Of Diagnostic Confirmation

Many conditions included in newborn screening programs are rare, and there may be poor access to diagnostic confirmation testing in the United States or even internationally. In such cases, it is more difficult to follow-up on cases with positive results, and the results provided by research laboratories may be more difficult to interpret and communicate to child health professionals and families than those from diagnostic laboratories. Furthermore, in the United States it may be ethically or legally problematic to report results from tests conducted by laboratories that are not certified by the Clinical Laboratory Improvement Amendments (CLIA). On the other hand, some conditions can be confirmed locally because of the wide availability and relative simplicity of the confirmatory test or service. Thus, different values were assigned based on the ease of diagnostic confirmation. One hundred points were allotted to this feature of a condition.

8. Acute Management

As with diagnostic confirmation, the availability of health care professionals who have expertise in the acute management of the condition could be limited. Thus, higher values were assigned to conditions for which acute disease management is readily available. One hundred points were allotted to this feature of a condition.

9. Simplicity Of Therapy

Therapeutic interventions range from highly specialized (e.g., bone marrow/umbilical cord blood transplantation) to extremely simple (e.g., vitamin supplementation, avoidance of fasting). A higher value was assigned to simpler therapies since simplicity determines whether infants requiring follow-up can be managed locally or whether subspecialist care is required. The acute management of many metabolic disorders often requires the involvement of metabolic disease physicians who are not readily available in many geographic locations. On the other hand, for example, aspects of CH may be managed by child health professionals, and when specialists are required, they are more widely available. Some conditions also might allow for greater levels of family involvement in treatment. One hundred points were allotted to this feature of a condition.

Collecting the data

One goal of the data collection process was to include a broadly representative group of participants. A second goal was to use a method that would allow quantification of expert opinion. In addition to data gleaned from the scientific literature, input and opinion were sought from a wide array of child health professionals, subspecialty care experts and individuals interested in newborn screening. Respondents were not anonymous, and were asked to select one or more of the following

categories to describe their personal and/or professional role(s) in relation to newborn screening:

1. Provider of screening services (TESTING)
2. Provider of screening services (FOLLOW-UP)
3. Provider of screening services (ADMINISTRATION)
4. Provider of screening services (POLICY)
5. Provider of diagnostic services
6. Child health professional
7. Specialty care provider
8. Consumer

As discussed previously, many criteria were perceived differently by these diverse constituencies. Distinguishing among respondents allowed the expert group to independently assess the views of these different groups.

For each condition, steps were taken to ensure that those asked to provide information and those who provided information were broadly representative of the interest groups involved. A large number of acknowledged experts for each condition and specific consumer and professional organizations were asked to provide input through multiple professional groups (e.g., the Society for Inherited Metabolic Disease (SIMD), ACMG). Individuals from public health and newborn screening programs were offered the opportunity to participate through listservs of their representative organizations. This included listservs managed by HRSA/MCHB, NNSGRC, the Association of Public Health Laboratories, and others. To ensure that the perspectives of consumers were available for consideration, consumers were reached through listservs of NNSGRC, Genetic Alliance, and others. To ensure that there were several scientific and clinical experts for each condition, specific individuals were identified from recent publications, disease support groups, and professional groups. In addition, the data collection instrument used was made widely available through the ACMG web site (www.acmg.net). Due to the large and overlapping numbers of individuals participating in these listservs, it is not possible to state the number of potential participants who were contacted. Geographic origin and role or interest in newborn screening of survey participants was monitored to ensure that respondents were broadly representative.

Respondents were given the opportunity to score each criterion or mark it as unknown “U,” an important option, because not all of those asked to participate were sufficiently familiar with the many aspects of all of the diseases for which responses were sought. However, the option also had implications for how the data were aggregated for analysis. The data were analyzed as means and medians for each criterion, as the average of total scores for each responder, and as sums of means and medians of all respondents to a particular criterion. After considering these different possibilities, it was decided that the results for any given condition would be expressed as the sum of the mean of the scores for each criterion. (The difficulty with using the sums of the means arises from different numbers of scorers, and scores varying in the comparisons, which obscures the distribution and confidence intervals of the final scores. The alternative approach using the sum of the

medians was not used as the primary statistic because it tends to minimize dissent from the consensus. In later figures, conditions are ordered around the sum of the means and medians are otherwise shown. However, as previously discussed, for purely objective criteria, the data as evidenced by the scientific literature was applied and included in the sums rather than the survey information.)

Developing and integrating the evidence base

In the second tier of the assessment, the evidence base for the conditions was established and an algorithm through which conditions were reassessed was developed. The quantification of expert opinion or scoring system then becomes part of a broader assessment of the scientific literature related to the conditions, tests, and treatments in the second level of the assessments. The evidence from the scientific literature, with supporting references for each criterion of each condition, was reviewed as shown in the fact sheets (Appendix 1). Evidence was derived from a systematic review of:

1. Clinical evidence;
2. Cost/economic evidence and modeling;
3. Reference lists obtained from PubMed and Medline;
4. Books;
5. Health technology assessments commissioned by the U.K. National Screening Committee;
6. The Internet, including disease-specific support groups; and
7. Professional guidelines.

Epidemiology studies, when available, were assessed for study design, the nature of the subjects and the outcomes that were measured, and the effectiveness of the treatment.

Statistical analysis of survey results allowed for a score to be assigned to each condition which determined its ranking and initial placement in one of three categories (high scoring, moderately scoring, and low scoring or lacking a newborn screening test). After the assignment of conditions to one of the three categories, the evidence base on the condition, as validated by acknowledged experts in the conditions in question, was used to determine if the conditions met critical criteria categories. Experts in specific conditions were identified by the Conditions and Criteria Work Group and included many individuals who had participated in the data collection process.

Several critical criteria were identified that reflected the priorities and principles of the expert group. These include:

1. The existence of a sensitive and specific test that has been validated in a large general population;
2. The availability of an efficacious treatment;
3. A determination that the natural history was sufficiently well understood to justify placement in a core panel of conditions;
4. Determination of whether a clinically significant condition not in the core panel would be identified because it is part of the differential diagnosis of a core panel condition; and

5. Whether a clinically significant condition would be revealed by a multiplex technology and whether it was part of the differential diagnosis of a core panel condition.
6. Further, it was recognized that some tests allow for the definitive identification of unaffected carriers, and that such results should be communicated to a responsible individual in the health care system.

The fact sheets for each condition were reviewed by at least two experts for each condition to validate the information and assign a level of quality to the evidence. Levels of evidence correspond to those defined by the AAP Steering Committee on Quality Improvement and Management³² as follows:

Level 1: Evidence is derived from well-designed randomized controlled trials or diagnostic studies on relevant populations.

Level 2: Evidence is derived from randomized controlled trials or diagnostic studies with minor limitations; overwhelming, consistent evidence from observational studies.

Level 3: Evidence is derived from observational studies (case control and cohort design).

Level 4: Evidence is derived from expert opinion, case reports, and reasoning from first principles.

The evidence was aggregated into four groups (the condition, the test, the diagnosis and the treatment) and a level of evidence quality was assigned to each group by the experts for each of the conditions. Each fact sheet includes the names of the experts who validated the data and the level of quality of the studies from which the evidence is derived.

C. Results

Responses were received from 289 individuals, many of whom represented more than a single interest group, for a total of 582 represented areas of interest. The majority of the survey information was provided by experts in the clinical and scientific aspects of the individual conditions. The regional distribution of responses and areas of expertise of the respondents from the United States are shown in Table 3. The table also

correlates the number of responses to the birth rate in each region (based on Census 2001 data). In the United States, no responses were received from the following States: Idaho, Kansas, Montana, North Dakota, South Dakota, West Virginia, and Wyoming. International responses were from Australia (4), Brazil (1), Canada (5), Chile (1), Croatia (1), Denmark (1), Finland (1), France (1), Germany (1), Italy (3), The Netherlands (1), Switzerland (1), and the United Kingdom. Most were from recognized experts in the field who were actively solicited by members of the working group for their input about specific conditions. At least three experts provided information on each condition.

Overall, a total of 3949 condition profiles were obtained. On average, seven conditions were scored per responder. Of the 84 conditions, 30 (36%) received more than 50 responses, and 5 (6%) < 20. The average number of profiles per condition was 47 ± 20 ; the range was 14-120. The corrected total for the 84 conditions was 3796; the number of responses for each condition is listed in Table 4. This table also shows the proportion of respondents who were unable to respond to one or more of the individual criteria and is reflected as "missing data" for each condition. This option was most frequently used in scoring criteria related to attributes of the screening test itself, with 11% of respondents not including all of the requested information.

Additional input, both domestic and international, was provided by individuals who were asked to discuss many of the broad issues under consideration by the work groups. The committee is particularly grateful for the assistance of Dr. Rodney Pollitt (Sheffield, UK), who provided insights into the system used in the United Kingdom; Dr. Adelbert Roscher (Munich, Germany), who provided insight into the recent newborn screening and MS/MS decision-making process undertaken by the German Democratic Republic; and Dr. Edwin Naylor (Pittsburgh, PA), who provided insight into the decision-making process of NeoGen Screening (now Pediatrix). In addition,

Table 3
Geographical distribution of respondent profiles

Region	Provider screening services					Specialty care provider							Total
	Testing	Follow-up	Administration	Policy	Consumers	Diagnostic services	Primary care	Endocrinology	Hematology	Inf. diseases	Genetics	Inborn Errors of Metabolism	
West	5	17	5	8	10	11	0	8	2	1	4	12	83
Midwest	8	23	4	16	14	20	1	5	2	1	12	18	124
Northeast	13	29	8	14	22	30	3	11	6	1	20	25	182
South	4	10	2	5	15	6	4	3	0	0	7	6	62
Southeast	2	6	2	6	22	9	1	5	3	0	7	6	69
Total US	32	85	21	49	83	76	9	32	13	3	50	63	520
International	11	11	5	5	0	15	1	0	3	0	0	9	60
Not provided	0	0	0	0	2	0	0	0	0	0	0	0	2
Total	43	96	26	54	85	91	10	32	16	3	50	72	582

Table 4
Survey scores of all conditions (sorted by score in descending order)

Condition	Code	Responses	Missing data (%)	Score (sum of the means)	Rank (%ile)
Medium-chain acyl-CoA dehydrogenase deficiency	MCAD	90	4	1799	1.00
Congenital hypothyroidism	CH	84	3	1718	0.99
Phenylketonuria	PKU	120	3	1663	0.98
Neonatal hyperbilirubinemia (Kernicterus)	HPRBIL	8	5	1584	0.96
Biotinidase deficiency	BIOT	68	2	1566	0.95
Sickle cell anemia (Hb SS disease)	Hb SS	55	8	1542	0.94
Congenital adrenal hyperplasia	CAH	93	7	1533	0.93
Isovaleric acidemia	IVA	53	3	1493	0.89
Very long-chain acyl-CoA dehydrogenase deficiency	VLCAD	58	2	1493	0.89
Maple syrup (urine) disease	MSUD	84	10	1493	0.89
Galactosemia	GALT	85	3	1473	0.88
Hb S/β-thalassemia	Hb S/βTh	43	8	1455	0.87
Hb S/C disease	Hb S/C	45	4	1453	0.86
Long-chain L-3-OH acyl-CoA dehydrogenase deficiency	LCHAD	58	3	1445	0.84
Glutaric acidemia type I	GA I	58	3	1435	0.83
3-hydroxy 3-methyl glutaric aciduria	HMG	28	4	1420	0.82
Trifunctional protein deficiency	TFP	42	5	1418	0.81
Multiple carboxylase deficiency	MCD	46	2	1386	0.80
Benign hyperphenylalaninemia	H-PHE	76	3	1365	0.78
Methylmalonic acidemia (mutase deficiency)	MUT	60	2	1358	0.77
Homocystinuria	HCY	80	2	1357	0.76
3-Methylcrotonyl-CoA carboxylase deficiency	3MCC	48	4	1355	0.75
Hearing loss	HEAR	45	4	1354	0.73
Methylmalonic acidemia (Cbl A,B)	Cbl A,B	46	2	1343	0.72
Propionic acidemia	PROP	68	2	1333	0.71
Carnitine uptake defect	CUD	46	2	1309	0.69
Galactokinase deficiency	GALK	47	7	1286	0.69
Glucose-6-phosphate dehydrogenase deficiency	G6PD	42	5	1286	0.67
β-Ketothiolase deficiency	βKT	33	6	1282	0.66
Citrullinemia	CIT	63	3	1266	0.65
Argininosuccinic acidemia	ASA	60	4	1263	0.64
Tyrosinemia type I	TYR I	68	4	1257	0.63
Short-chain acyl-CoA dehydrogenase deficiency	SCAD	51	7	1252	0.61
Tyrosinemia type II	TYR II	57	3	1249	0.60
Glutaric acidemia type II	GA2	52	4	1224	0.59
Medium/short-chain L-3-OH acyl-CoA dehydrogenase deficiency	M/SCHAD	21	11	1223	0.58
Cystic fibrosis	CF	65	12	1200	0.57
Variant Hb-pathies (including Hb E)	Var Hb	41	3	1199	0.55
Human HIV infection	HIV	29	8	1193	0.54
Defects of bipterin cofactor biosynthesis	BIOPT (BS)	60	3	1174	0.53

(continued)

Table 4
Continued

Condition	Code	Responses	Missing data (%)	Score (sum of the means)	Rank (%ile)
Medium-chain ketoacyl-CoA thiolase deficiency	MCKAT	23	13	1170	0.52
Carnitine palmitoyltransferase II deficiency	CPT II	45	5	1169	0.51
Methylmalonic acidemia (Cbl C,D)	Cbl C,D	45	4	1166	0.49
Argininemia	ARG	54	5	1151	0.48
Tyrosinemia type III	TYR III	42	5	1149	0.47
Defects of bipterin cofactor regeneration	BIOPT (Reg)	58	5	1146	0.46
Malonic acidemia	MAL	22	5	1143	0.45
Carnitine: acylcarnitine translocase deficiency	CACT	38	5	1141	0.43
Isobutyryl-CoA dehydrogenase deficiency	IBG	28	7	1134	0.42
2-Methyl 3-hydroxy butyric aciduria	2M3HBA	18	3	1132	0.41
Carnitine palmitoyltransferase IA deficiency (liver)	CPT IA	40	4	1131	0.40
2-Methylbutyryl-CoA dehydrogenase deficiency	2MBG	27	18	1124	0.39
Hypermethioninemia	MET	45	3	1121	0.37
Dienoyl-CoA reductase deficiency	DE RED	18	11	1119	0.36
Galactose epimerase deficiency	GALE	38	7	1066	0.35
3-Methylglutaconic aciduria	3MGA	21	5	1057	0.34
Severe combined immunodeficiency	SCID	69	6	1047	0.33
Congenital toxoplasmosis	TOXO	28	12	1041	0.31
Familial hypercholesterolemia (heterozygote)	FHC	25	2	1038	0.30
Carnitine palmitoyltransferase IB deficiency (muscle)	CPT IB	28	4	1009	0.29
Citrullinemia type II	CIT II	38	2	1001	0.28
Ornithine transcarbamylase deficiency	OTC	64	7	942	0.27
Guanidinoacetate methyltransferase deficiency	GAMT	23	1	922	0.24
Wilson disease	WD	25	4	922	0.24
Diabetes mellitus, insulin dependent	IDDM	51	16	891	0.23
Neuroblastoma	NB	14	4	864	0.22
Arginine: glycine amidinotransferase deficiency	AGAT	21	2	861	0.20
Turner syndrome	TURNER	36	4	847	0.19
Adenosine deaminase deficiency	ADA	20	4	841	0.18
Carbamylphosphate synthetase deficiency	CPS	55	2	833	0.17
Alpha 1-antitrypsin deficiency	A1AT	18	12	819	0.16
Congenital cytomegalovirus infection	CMV	18	12	779	0.14
Duchenne and Becker muscular dystrophy	DMD	29	3	776	0.12
Fragile X syndrome	FX	35	4	776	0.12
Congenital disorder of glycosylation type Ib	CDG Ib	34	5	766	0.11
Smith-Lemli-Opitz syndrome	SLO	45	3	759	0.10
Biliary atresia	BIL	15	4	744	0.08
Hurler-Scheie disease	MPS-1H	48	7	707	0.07
X-linked adrenoleukodystrophy	ALD	38	2	705	0.06
Fabry disease	FABRY	46	6	661	0.05

(continued)

Table 4
Continued

Condition	Code	Responses	Missing data (%)	Score (sum of the means)	Rank (%ile)
Lysosomal storage diseases	LSD	38	8	638	0.02
Creatine transport defect	CR TRANS	20	0	646	0.04
Pompe disease	POMPE	46	7	613	0.01
Krabbe disease	KRABBE	44	9	447	0.00

NOTE: Figure 5 shows the scores for all conditions that were evaluated, separated into groups based on the testing platforms (MS/MS for metabolic diseases, IEF or HPLC, for hemoglobinopathies, and all others).

several opportunities were offered for public comment over the course of these deliberations.

Based on responses to an independent survey that inquired as to the appropriateness of the criteria and the weights assigned, the expert group adjusted the scores assigned to some of the criteria. In particular, ambiguous language was clarified and a greater weight was assigned to the benefit of treatment to the infant. Scores for the parameters of the screening tests were increased to recognize the inherent value of multiplex technologies to public health.

Figures 1 and 2 display the raw data for MCAD and PKU, which were selected as representative conditions for demonstrating how the data collected for individual criteria are charted and aggregated to reach the final scores. Each respondent is listed over columns and the score offered for each criterion is shown. The sums of the mean and median scores are shown. Figures 3a through 3e display side-by-side summary data for each of the criteria used to evaluate the conditions with MCAD on the left and PKU on the right. In the top panel, the total score for each respondent is shown. The remaining panels show the scores for 18 of the 19 individual criteria (the availability of test criterion is not included) used to evaluate the conditions. The complete data in tabular form are displayed in Table 4, in which the scores are reflected as sums of the means for all conditions. The number of respondents for each condition is shown. The sums of the mean scores for all of the conditions evaluated, regardless of whether a screening test is available, are shown in Figure 4, Figure 5.

Figure 6 separates those conditions that have an acceptable, validated, population-based screening test from those lacking a test. The left side of the graph shows the conditions that have an adequate screening test currently available, while those shown on the right side lack a screening test. Among the conditions with a test, MCAD deficiency, CH, and PKU score the highest in this analysis, followed by BIOT, sickle cell anemia, CAH, isovaleric acidemia, VLCAD deficiency, MSUD, GALT, hemoglobin S/ β -thal disease, hemoglobin SC disease, LCHAD deficiency, glutaric acidemia type 1, and HMG. Conditions without a test are included because they reflect the need to focus on particular aspects of the disease in order for it to be considered for newborn screening.

D. Discussion

A number of considerations influenced the final decisions regarding which conditions should be included in a core screening panel. As previously discussed, using a two-step process, the information gathered with the data collection instrument and the review of the scientific literature provided information used to assign a score for each condition. This approach also allowed for those conditions with screening tests that have been validated in general populations to be distinguished from those conditions for which a population-based validated test was not available. The scores were first used to make some general decisions based on the highest scoring conditions. In particular, the inclusion of several conditions that are screened by either IEF or HPLC (hemoglobinopathies) and MS/MS (acylcarnitines and fatty acid oxidation disorders) led the expert group to make decisions regarding multiplex technologies and how the results should be handled. Once the conditions were separated into groups defined by either the individual condition or by the multiplex test that detects many conditions, the scoring system could be overlaid to see how conditions compare to one another within these groupings, or in total.

Defining and counting the conditions

Careful consideration of several factors is required to answer the seemingly basic question of how many conditions should be screened for in a newborn screening program and how they should be defined. These factors include: 1) the clinical, biochemical, and molecular complexity of the conditions under consideration; 2) the progress constantly made in our understanding of their natural history and etiology; 3) the impact of implementing multiplex platforms that allow the simultaneous detection of numerous biochemical markers; and 4) the gaps that appear to exist in the level of clinical knowledge among stakeholders involved with, or advocating for, the decision to pursue ever greater numbers of conditions. Indeed, counting has become increasingly problematic to the point that a competition seems to be taking place in which the apparent superiority of a newborn screening program or private laboratory is staked on the sole basis of quantity, with disproportionate consideration given to quality. This concept has caught the attention of the media that constantly tell the pub-

lic-at-large that the more conditions that are screened in a particular State, the better that program must be. As a direct consequence of this behavior, the number of conditions is perceived by the public and policy-makers as a scorecard, often leading to either inflated or inaccurate figures. For example, 22 States offering screening by MS/MS have included LCHAD deficiency in their panels, yet only half of the same programs claim to be screening for trifunctional protein deficiency, perhaps being unaware that the biochemical phenotype in blood-spots is essentially identical between the two conditions. Thus, the context in which screening is “quantitated” must be standardized.

This situation is not a new development brought on by modern technologies. Since the beginning of PKU screening, this has been a complex issue. The screening method for PKU led to follow-up testing to separate the patients with tyrosinemia and/or bipterin defects. Thus, many programs included tyrosine in their screened conditions, and considered bipterin defects as merely an anomaly of PKU screening that should be combined with PKU and given an asterisk when counting the number of PKU cases detected. This is hardly satisfactory when questions are asked about the incidence of the secondary targets or the outcomes of those subtypes.

When screening for sickle cell anemia became an important addition to screening panels, the singular condition of SS disease was usually counted even though the testing methodologies used could detect many different clinically significant hemoglobinopathies. Screening for sickle cell anemia progressed to screening for sickle cell diseases (SC and S β -thal) but this screening was still counted as screening for a single disorder with many other conditions detected secondarily. Further, although these are the three primary targets of hemoglobinopathy screening, the methodologies of IEF or HPLC employed in hemoglobinopathy screening can reveal over 700 variant hemoglobins, of which about 25 are considered to be of clinical significance and are reported out by some screening laboratories. Some States may only report SS dis-

ease, some SS, SC and S β -thal, and others a variable number of the other clinically significant variants. Hence, just for this one group of conditions, one can argue that a program that reports out 28 of these variants actually screens for 28 conditions. For a test involving a functional endpoint such as severe hearing loss, there are a large number of “conditions” for which the test screens.³³ There are over 77 loci for nonsyndromal hearing loss conditions, 31 loci for syndromal hearing loss conditions, as well as some of the “environmental” causes of hearing loss that would be amenable to DNA-based testing such as presence of the cytomegalovirus or other infectious agent genomes. Hence, what is considered a single condition screen, congenital hearing loss, may be considered a screen for at least 108 individual conditions at the etiologic level.

If one takes the set of conditions included in both the proposed core panel and secondary target groups, each entity reflects the significance given to a spectrum of possible criteria. In the proceedings of the working group charged with this task, choices were made to strike the best compromise between established practices, the expert opinions, and scientific evidence. In reality, counting could have been very different if this had been approached in a pragmatic way using any of the following criteria:

1. Phenotype of the condition;
2. Established groups of conditions (e.g., organic acidurias, hyperphenylalaninemias);
3. Primary marker (e.g., tyrosine, C8 acylcarnitines);
4. Test (e.g., MS/MS, IEF);
5. Response to treatment (e.g., responsiveness to cofactors, vitamins); and
6. Number of loci linked to a common phenotype (e.g., hearing loss genes as discussed above).

Table 5 shows how different “counting” could be if the criteria above were applied independently. For instance, hearing loss is a single phenotype of one group of conditions for which the primary marker is hearing loss that is detected by one test-

Table 5
Discrepancies in counting conditions using different criteria

Counting conditions according to	CORE PANEL	(NOT included if overlapping with core panel) SECONDARY TARGETS	TOTAL
Clinical phenotype (1)	27	14	42
Established groups of conditions (2)	10	0	10
Primary marker (3)	22	29	51
Test platform (4)	9	2	11
Response to treatment (5)	32	14	46
Number of loci (6)	142	28	170
Expert group (7)	29	25	54

(1) All clinical subsets (e.g., severe, mild) considered as a single entity.

(2) Organic acids disorders, hemoglobinopathies, endocrine disorders.

(3) Analyte with best sensitivity and specificity (e.g., C8 for MCAD or phenylalanine for the hyperphenylalaninemias).

(4) Either singleton test or multiplex platform count as one.

(5) Significant in a few cases (e.g., responsive versus non-responsive forms to a particular treatment).

(6) Based on OMIM (), with modifications.

(7) Selected from a total of 84 conditions.

ing platform, audiometry. The single response to treatment for the group is improved hearing or communication. However, as previously discussed, there are at least 108 genes for conditions associated with hearing loss. Similarly, while C8 is a primary marker of MCAD, it's also a primary marker for GA-II, M/SCHAD and MCKAT. It is detected in a single multiplex platform, MS/MS. Treatments are similar but as indicated above, and multiple conditions are associated with the marker.

It is evident that quantitation and categorization of newborn screening disorders remains imperfect and inconsistent and that, until standardized, there will continue to be confusion about the extent of screening in individual programs and the nation. The expert panel recognizes these disparities and their rationale, and recommends the implementation of a standardized and common nomenclature for an objective and scientifically sound description of the screening test panel being offered and the reporting of results. Such a classification system would require some consensus among the newborn screening and subspecialty communities, but should be possible. Standardization of panels, and consistent screening methods and case definitions will allow more pooling of available data on the utility of screening.

Integrating the evidence base with the survey results

Information obtained from the scientific literature and the surveys was used to create the fact sheets that were developed for each condition (see Appendix 1). The fact sheets are structured to provide summary information describing:

1. The type of condition;
2. The test;
3. The extent to which United States newborns are being screened for the condition;
4. Whether there is apparent ethnic variability in incidence;
5. The number of individuals providing information on the condition;
6. The proportion of scores from survey respondents considered valid; and
7. Citations in PubMed as of February 2004.

Information obtained from the surveys is shown on the left side of the first page. The percent of maximum score of the survey respondents is shown next to each criterion. The data from the two criteria for which there was the lowest correlation among respondents is also shown on the left side of page 1. The evidence from the literature is shown on the right side of the first page. Additional summary information including the scores (maximum of 2,100) is shown along with an assessment of whether the data from the surveys are consistent with the evidence from literature. Significant discrepancies are discussed in the comment box. Although the language of the criterion is often not identical to that expressed in the literature, there was significant correlation between the survey results and the evidence from the literature. The fact sheets for all other conditions evaluated are provided in Appendix 1.

Influence of testing technology

New technology has been one of the driving forces in the evolution of newborn screening programs in the United States and is a critical factor in the evaluation of a condition to determine how appropriate for screening it is. Typically, determining the appropriateness of newborn screening was based on the conditions themselves and their associated testing methods. However, new technologies often raise questions that have not yet been addressed. Multiplex methods such as genomic arrays require that the sequence tested deliberately be placed in the array. This is distinct from technologies that look globally at a class of molecules, for example, IEF or HPLC that reveal all hemoglobin variants, or an MS/MS run to detect acylcarnitines that reveal compounds in the C2 through C18 range. Complicating the use of MS/MS is the fact that many of the compounds identified are associated with more than one condition and these conditions may not have similar clinical and laboratory features. Thus, the criteria used to judge whether to include a condition in a newborn screening panel will vary among the conditions. It becomes difficult to compare a condition that has a unique test/technology that tests only for the condition of interest to a technology that can detect many conditions, some of which are related through their differential diagnosis, while others involve independent compounds in the MS/MS profile. The use of MS/MS for acylcarnitines, for example, differs from its use for detection of amino acid disorders in which there is little overlap between the analytes associated with the conditions. Table 6 shows the relationships between analytes for high scoring conditions and those of lower scoring conditions.

Independent decisions were made about conditions screened using MS/MS and HPLC or IEF for hemoglobinopathies. One reason is that among the acylcarnitine disorders there is little differentiation between the highest and lowest scoring conditions. For many conditions, the difference is accounted for by differing incidence figures—a criterion that loses some of its importance when the test for the more common conditions also can detect less common conditions.

It is important to note that two approaches are currently being used in screening with MS/MS. A majority of screening laboratories now run full profiles that allow them to visualize the full range of acylcarnitines or amino acid compounds. However, a minority operate their systems in a selective reaction monitoring (SRM) mode, which allows them to obtain results only on the subset of compounds that are associated with those conditions that are being targeted in the screening programs. Some programs use a combination of SRM and profiling with either approach, the screening test is driven more by analytes than by the conditions with which they are associated. An assessment of the advantages and disadvantages of the test results for each approach led to an expert group preference for the full-profile approach for four reasons.

First, in reviewing those acylcarnitine-associated conditions that were high scoring in this analysis (MCAD, IVA, VLCAD, LCHAD, GA1, HMG and TFP) (see Table 4), it was apparent

Table 6

Differential diagnosis between core panel and secondary target conditions

PRIMARY TARGETS		SECONDARY TARGETS
Higher Scoring	Lower Scoring	
MCAD		GA2 M/SCHAD MCKAT
PKU		H-PHE BIOPT (BS) BIOPT (REG)
Hb SS	Hb S/β-Th Hb S/C	VAR Hb
IVA		2MBG
VLCAD	LCHAD TFP	CPT II CACT
GALT		GALK GALE
BIOT (*)	MCD PROP	
MUT	Cbl A,B	Cbl C,D
HCY		MET
HMG	3MCC BKT	2M3HBA 3MGA
CUD		CPT IA
CIT	ASA	CIT II
TYR I		TYR II TYR III

NOTE: Codes are as listed in Table 2. A differential diagnosis is required between conditions listed in the same row. (*) indicates that biotinidase deficiency is occasionally diagnosed by MS/MS.

that several acylcarnitines must be analyzed in order to maximize assay specificity and sensitivity. A majority of the remaining conditions detected by MS/MS were also included in the differential diagnoses of the higher scoring conditions. Thus, screening for a core set of conditions ultimately results in screening for a much wider range of conditions.

Second, the use of MS/MS profiles allows for the maximal use of the technology for the identification of clinically significant conditions.

Third, the use of MS/MS profiles offers better quality control of preanalytic and analytic aspects of testing. Allowing all information to be assessed can reveal the presence of spurious signals and/or contaminants in the specimens or reagents and devices used in the test system.

Fourth, the use of MS/MS profiles enhances clinical interpretation of results by revealing anomalies in associated compounds or in compounds that provide internal standards against which excesses or deficiencies can be better interpreted. Hence, the expert group recommends that a full MS/MS profile should be analyzed, and any clinically significant results should be reported by the laboratory to the health care provider and family of the infant. Some of the conditions detectable by acylcarnitine profiling may turn out to be benign in a

number of cases (i.e., SCAD, 2MBCAD, and 3MCC). The secondary conditions detectable by a multiplex technology such as MS/MS or HPLC and included in a differential diagnosis for the primary target conditions can be screened at minimal additional cost and are, in fact, determined in the diagnostic setting during follow-up. There could be additional cost associated with diagnosis and follow-up, although many of these cases would be detected clinically after birth and higher costs would inevitably be incurred by the health care system and the family, although not as a result of the newborn screening program.

The expert group also devoted considerable discussion to the question of how best to present the results of analyses of conditions. As previously discussed, the lists of conditions used are inherently longer than the lists many States use to describe the newborn screening tests they offer because the expert group chose to break down the heterogeneity of conditions by listing them by etiologic type or by the analytes associated with the conditions. It would be inappropriate to consider this list of conditions as a scorecard for the number of conditions screened. It is only by considering each condition in each of its etiologic forms that a direct analysis can be done.

In the following section, diseases are assigned to categories as a means of conducting the analyses (see Tables 7 and 8). The main category, referred to as the core panel, includes those conditions considered appropriate for newborn screening. The 29 conditions in this core panel are similar in that they all have:

1. Specific and sensitive screening tests;
2. A sufficiently well understood natural history; and
3. Available and efficacious treatments.

Table 7
The core condition panel

MS/MS				
Acylcarnitines		Amino acids		
9 OA	5 FAO	6 AA	3 Hb Pathies	6 Others
CORE PANEL				
IVA	MCAD	PKU	Hb SS*	CH
GAI	VLCAD	MSUD	Hb S/β-Th*	BIOT
HMG	LCHAD	HCY*	Hb S/C*	CAH*
MCD	TFP	CIT		GALT
MUT*	CUD	ASA		HEAR
3MCC*		TYR I*		CF
Cbl A,B*				
PROP				
BKT				

Codes are as listed in Table 4. OA, disorders of organic acid metabolism; FAO, disorders of fatty acid metabolism; AA, disorders of amino acid metabolism; Hb Pathies, hemoglobinopathies. (*) See individual condition discussions.

Table 8
The secondary target condition panel

SECONDARY TARGETS				
6 OA	8 FAO	8 AA	1 Hb Pathies	2 Others
Cbl C,D*	SCAD	HYPER-PHE	Var Hb*	GALK*
MAL	GA2	TYR II		GALE
IBG	M/SCHAD	BIOPT (BS)		
2M3HBA	MCKAT	ARG		
2MBG	CPT II	TYR III		
3MGA	CACT	BIOPT (REG)		
	CPT IA	MET		
	DE RED	CIT II		

Codes are as listed in Table 4. OA, disorders of organic acid metabolism; FAO, disorders of fatty acid metabolism; AA, disorders of amino acid metabolism; Hb Pathies, hemoglobinopathies. (*) Identifies conditions for which specific discussions of unique issues are found in the main report.

The expert group concluded that conditions with evidence-validated scores equal to or above 1,200 meet these key criteria and should be considered appropriate for newborn screening.

Analysis of the distribution of scores among the conditions in Figure 7 shows that around a score of 1,250, one moves into a group of conditions that are part of the differential diagnosis of higher scoring conditions, but for which natural history is less well understood or efficacious treatment is lacking. These conditions occupy the middle third of the curve. CF (1,200) is the only condition currently screened that scores in this range but is not part of the differential diagnosis of a higher scoring condition. (Its lower score may reflect the ongoing debate about the benefits of screening for CF, despite the evidence for screening and the lack of evidence of significant harms from screening.)^{34–35} Otherwise, all conditions in this middle third scoring between tyrosinemia type I (score = 1,257; 63rd centile) and galactose epimerase deficiency (score = 1,066; 35th centile) are part of the differential diagnosis of another higher scoring condition. The expert group recognizes that it is difficult to draw a line in a continuum that would reasonably discriminate between groups of conditions. Programs should appreciate that scoring cut-offs may have wide and varying confidence limits due to differences in numbers of responders. The final scores represent a rough relative approximation of ranking of disorders and serve only as an initial step to guide decision-making; analysis of the evidence base for the score needs to be included in the decision-making process.

Conditions then were redistributed between the core panel and the secondary target category on the basis of the evidence related to the availability of an efficacious treatment and a well understood natural history. Other conditions were moved from the “not appropriate for newborn screening category” to secondary targets if they were revealed by the multiplex technology used to identify core panel conditions. SCAD, IBG, ARG and DE RED were moved into the secondary target category on this basis. Among conditions initially placed in the core panel category on the basis of the survey score, CPT-II was shifted to the secondary target category on the basis of the lack

of a proven efficacious treatment. Several conditions were moved to the secondary target category on the basis of scientific evidence indicating that the natural history was not sufficiently well understood. These include TYR-II, GA-2, and M/SCHAD. GALK deficiency was moved to the secondary target category on the basis of the relatively limited burden of disease and the fact that a second test is usually required to screen for the condition. G6PD was moved to the category of conditions not recommended for newborn screening because of a limited knowledge of the natural history of the mutations in the G6PD gene found in the United States. There is also limited knowledge of the implications of these mutations with regard to development of severe hemolytic disease in the United States population. Additionally, because G6PD is not identified in the course of screening for other core conditions, it was not placed in the secondary target category. Finally, a subset of conditions was identified for which carrier status could be established on the basis of the screening test result and for which reporting is considered appropriate. These include MCAD, VLCAD, Hb-pathies, 3MCC, CUD, and CF.

The next group of conditions includes those that are clinically significant and are part of the differential diagnosis of a condition listed in the core panel or that are revealed through a multiplex technology. Note that secondary hemoglobinopathies are revealed in the screening laboratory while most others are revealed in the diagnostic setting during follow-up. Table 8 lists the conditions in this secondary category. Table 5 shows the relationships among many of the core conditions and the conditions included in their differential diagnoses (or secondary targets). In particular, some of the metabolic conditions in this group are characterized by having a sensitive and specific test, but a deficiency in the availability of an efficacious treatment or limited knowledge of the natural history of the condition, although there may be sufficient knowledge to justify the reporting of test results to the family and health care provider of the infant.

The recommendation to report all clinically significant results is an approach similar to that taken for hemoglobinopa-

thy screening, in which a core set of conditions is screened. The technologies of choice in many laboratories for hemoglobinopathy screening are IEF and HPLC, which can detect the full range of more than 700 hemoglobin variants, including those in the core panel, for which clinically significant variants are reported.³⁶ By handling hemoglobinopathies in a way similar to the acylcarnitine and amino acid disorders screened for by MS/MS, the expert group was left with a much smaller group of conditions to consider independently for screening suitability. These conditions have adequate screening tests and efficacious treatments, but they are detected by methods other than MS/MS, and usually as singleton tests.

Table 9 lists the conditions that were determined to be without a screening methodology that has been adequately validated for general population-based screening. Kernicterus risk as determined by the identification of hyperbilirubinemia stands out in this group as being a very high scoring condition.

Figure 8 shows the distribution of conditions into the: core panel (29 conditions); secondary target category (25 conditions); no test available (23 conditions), those excluded from

Table 9

Conditions for which Newborn Screening is NOT Indicated at this Time

MS/MS				
Acylcarnitines		Amino acids		
OA	FAO	AA	Hb Pathies	Others
No Test				
	CPT-1B	OTC	HPRLBIL	FX
		CPS	FHC	CDG-1b
			SCID	SLO
			IDDM	ALD
			GAMT	MPS-1H
			WD	FABRY
			AGAT	CR TRANS
			NB	LSD
			TURNER	POMPE
			BIL	KRABBE
Excluded				
			ADA	
			A1AT	
			DMD	
			G6PD*	
Deferred				
			HIV	
			TOXO	
			CMV	

Codes are as listed in Table 4. OA, disorders of organic acid metabolism; FAO, disorders of fatty acid metabolism; AA, disorders of amino acid metabolism; Hb Pathies, hemoglobinopathies. (*) Identifies conditions for which specific discussions of unique issues are found in the main report.

newborn screening categories due to other inadequacies in meeting the criteria (4 conditions), and the three conditions on which we deferred decision-making.

Selected condition discussions

The following conditions represent a group for which there was either deviation from the adopted data processing plan or for which unusual issues justify additional discussion. It is important to realize that the data on the laboratory sensitivity and specificity of many conditions identified by MS/MS is suboptimal, though it was sufficient to lead the expert group to classify them as it has done.

Congenital Adrenal Hyperplasia (CAH)

Table 7 CAH includes a number of forms of the disease. The most common is 21 hydroxylase (21-OH) deficiency, which accounts for 95% of cases and is the general form that has been considered. The primary marker used in newborn screening for 21-OH, 17-hydroxyprogesterone (17-OHP), is most sensitive in identifying infants with the severe salt-wasting form in which elevations are very high. The degree to which 17-OHP is elevated in the nonsalt-wasting forms is variable. Hence, sensitivity in detecting this form by newborn screening is reduced. The 21-OH forms of CAH were not subdivided as were the hyperphenylalaninemias because the forms of 21-OH are caused by the same gene. However, many programs consider the identification of newborns with the nonsalt-wasting form to be a by-product of screening for the primary target, the salt-wasting form. In the salt-wasting form, most virilized females should be clinically detectable because of "ambiguous genitalia" or as virilized females. However, it is important to identify the males by screening to prevent early morbidity and mortality. The other CAH types found in the remaining 5% of patients are not detectable generally by current screening strategies.

Galactokinase Deficiency (GALK)

Table 8 Galactokinase deficiency scored 1,286 points in the analysis. However, the only consistent phenotype is cataracts. Further, in order to screen for GALK, an additional test is required. Most screening laboratories include a combination of the Beutler fluorescent spot screening test and a fluorometric or bacterial inhibition assay for total galactose. Because GALK is very rare and is part of the differential diagnosis of GALT, it has been designated as a secondary target.

Glucose 6-Phosphate Dehydrogenase Deficiency (G6PD)

Table 9 G6PD deficiency is included in newborn screening programs in some countries, particularly in Asia and the Mediterranean, where it is the most common enzymopathy. Newborn screening programs in the Philippines and in Taiwan have reported incidence figures of 1 in 65. In the United States, G6PD screening is provided as part of the screening panel for the District of Columbia – the only program to mandate and provide screening for G6PD deficiency (Missouri has mandated G6PD screening but has not yet implemented the screen-

ing). The vast majority of the clinical data are from countries in which the risk factors (e.g., ingestion of fava beans, infections, and drugs such as sulfonamides and antimalarials) associated with G6PD status are more common and in which the prevalence is higher (e.g., tropical Africa, Middle East, tropical and subtropical Asia and in some areas of the Mediterranean). There is very limited data available from any screening program in the United States, and the opinion of hematology experts is that the variants that exist in the United States African American population are clinically benign unless the individual is in a severely compromised (i.e., oxidized) state, usually resulting from drug exposure. Additional data are needed from programs now screening for G6PD before this condition can reasonably be considered for inclusion in a mandated core panel of screening conditions. Programs currently screening for G6PD are encouraged to collect and publish the data for determining clinical relevancy and analytical specificity and sensitivity of tests being used. Further, and as discussed below in the context of hyperbilirubinemia, some conditions are not mutually exclusive. Appropriate monitoring and management of jaundice could identify those cases at risk for Kernicterus or biliary atresia.

Hemoglobinopathies (Hb Pathies)

Table 8 Hemoglobinopathies are screened by HPLC or IEF in most programs. The primary focus of the review of scientific literature was on sickling disorders, since they have been the primary targets of newborn screening. However, there are over 700 hemoglobin variants identified by the methods used for screening, and 25-30 are considered clinically significant. Many of these conditions are associated with an Hb SS allele, but not all. Among these variant hemoglobinopathies, Hb E is by far the most common. The expert group agreed with the current recommendations that all clinically significant hemoglobinopathy variants be reported to health care professionals. It is appreciated that there may be conditions that occur more commonly in subpopulations, such as the case of Hb E in the Hmong population, and that may alter local screening practices.

Homocystinuria (HCY)

Table 7 Homocystinuria is screened for by detection of an elevated concentration of methionine, a secondary biochemical marker of the condition. The differential diagnosis of HCY includes other defects of methionine metabolism, unrelated liver disease, common dietary artifacts (total parenteral nutrition), and analytical issues (lability of methionine internal standard).³⁷ Hence, screening for HCY has a lower sensitivity than other amino acid disorders included in the core panel, and requires special attention in result interpretation to minimize the rate of false positive results. Although a primary screening based on methionine is less than ideal, the identification of newborns with a potentially treatable condition was a determining factor for the high score assigned to HCY in the survey and its inclusion in the core panel. This situation is likely to evolve when a second tier test capable of measuring

total homocysteine in bloodspots becomes routinely available by MS/MS or other methods; an improvement that will strengthen the inclusion of HCY in the core panel.

Hyperbilirubinemia (HPRLBIL)

Table 9 Based on the responses of seven experts asked to complete the data collection instrument, this was among the highest scoring conditions. However, the expert group determined that there was not a screening methodology that was sufficiently well validated in a large newborn population to justify mandated universal screening at this time. Although bilirubin test result nomograms have been validated in smaller studies, the current nomograms are not sufficiently reflective of the broad population. There are also risk factors for hyperbilirubinemia associated with other conditions such as G6PD deficiency that are assessed independently. Additionally, in order for bilirubin to be used as a marker of this condition, a specimen would have to be taken and testing would likely have to occur in the local nursery, because results would need to be rapidly available based on current understanding of hyperbilirubinemia. Therefore, the question is raised whether this should be a mandated newborn screen or, rather, be instituted as an appropriate standard medical practice for any newborn.³⁸ Currently, universal testing for hyperbilirubinemia is not routinely conducted in most hospitals.

Methylmalonic Acidemia

Methylmalonic acidemia (MMA) exists in several etiologic forms caused by defects of either the apoenzyme (MMA-CoA mutase) or the biosynthesis of the coenzyme (adenosyl-cobalamin). The forms associated with a coenzyme defect may overlap biochemically with acquired dietary deficiencies. The biochemical marker of MMA is propionylcarnitine. Overall, there is credible evidence of less than ideal sensitivity with the current testing technology (affected cases with normal concentration when tested at birth) and specificity (relatively high rate of false-positive results, including cases with relatively high levels that are followed up by perfectly normal plasma acylcarnitine and urine organic acid profiles). It is likely that the introduction of a second-tier test capable of measuring methylmalonic acid in bloodspots could improve the sensitivity and specificity of newborn screening for MMA and reinforce the inclusion of this condition in the core panel. Because newborn screening is considered a program that extends beyond the screening test itself, it was decided that the disorders characterized by an elevated propionylcarnitine (mutase deficiency, cobalamin A, B, C, and D deficiencies, as well propionic acidemia) should be subdivided, particularly since they have quite different natural histories and treatment options.

3-Methylcrotonyl-CoA Carboxylase Deficiency (3MCC)

Table 7 The natural history of 3MCC has been driven by the clinical ascertainment of patients presenting with severe acute episodes. However, since newborn screening with MS/MS began, several individuals have been identified with the analytes associated with the condition but without apparent clinical

manifestations. This situation includes cases where the abnormal metabolites found in the neonatal bloodspot were of maternal origin, subjects who are usually biochemically affected but symptom-free. All elements being considered, it is in the best interest of newborns affected with 3MCC that the condition be identified in all cases. 3MCC was therefore included in the core screening panel with the expectation that long term follow-up will lead to a better understanding of this condition and its clinical significance.

Tyrosinemia Type I (TYR I)

Table 7 TYR I is a condition caused by fumarylacetoacetate hydrolase deficiency that presents with severe liver and renal disease and peripheral nerve damage. If left untreated, most patients die of liver failure in the first years of life. Treatment with the drug NTBC (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione), diet, and liver transplant are now considered to be very effective. Newborn screening is based on the detection of an elevated concentration of tyrosine. There is evidence of less than ideal sensitivity with the current testing technology (affected cases with normal concentration when tested at birth) and poor specificity (very high rate of false positive results, mostly premature babies and newborns with liver disease of variable etiology). Although the introduction of a second-tier test capable of measuring succinylacetone in bloodspots could improve the sensitivity and specificity of newborn screening for TYR-I, the question of whether affected but asymptomatic newborns are being identified with any degree of consistency remains to be answered. It is a general and accepted concern that hepatorenal tyrosinemia may not be detected by MS/MS analysis of tyrosine concentration alone. However, TYR-I is included in the core panel for historical reasons and because of the effectiveness of treatment. It remains important not to exclude the diagnosis of tyrosinemia on the basis of a screen negative result.

Limitations of methodology

Over the course of this project a number of limitations became apparent. Conditions with limited available evidence reported in the scientific literature were more difficult to score and place in one of the three categories. Some conditions had been reported in 10 or fewer families in the world, and for other conditions, there were gaps in the evidence base in the literature. Many conditions were found to occur in multiple forms distinguished by age-of-onset, severity, or other features. In most cases, decisions related to newborn screening were based on the more severe and treatable forms of the conditions.

The knowledge base about genetic diseases grows through a common pathway and, unless a condition was already included in newborn screening programs, there was a potential for bias in the information related to some criteria. The most severe forms of genetic diseases are usually those first noted. As one moves into the families of these probands, this bias toward severity is reduced. However, it is not until a large general population has been studied that the true performance char-

acteristics of the various screening tests are appreciated. Because many of the conditions under consideration are very rare and the genetic etiologies may vary by ethnicity and other parameters, a population of considerable size is required to acquire a broad understanding of the condition.

Due to the aforementioned limitations, expert opinion that considered reasoning from first principles and the quality of the studies underlying the data contributed significantly to the placement of the conditions into particular categories.

Numerous barriers to implementing an optimal screening and follow-up program were identified. Recommended actions to overcome these barriers include the establishment of a national role in scientific evaluation of conditions and the technologies by which they are screened, standardization of case definitions and reporting procedures, enhanced oversight of hospital-based screening activities, long-term data collection and surveillance, and consideration of the financial needs of programs to allow them to deliver the appropriate services to the screened population.

Finally, there were limitations in both time and resources available to accomplish a project as broad and comprehensive as this. A large number of conditions commonly managed by differing subspecialists were assessed and, due to their rarity, it was not unusual that there may only be a handful of acknowledged experts of particular conditions in the world. It was also necessary to include a significant number of experts not directly involved in the expert group or its work groups. In order to broaden the number of individuals from whom we might draw for assistance with data collection and validation, it was necessary to consult with international experts.

In many ways, the analyses done under this project provide a current snapshot of the knowledge base from which recommendations are drawn. Decisions were made as to the adequacy of the evidence on which the recommendations are based. However, as is common for rare diseases, the acquisition of new knowledge is ongoing and long-term surveillance is needed to ensure that the evidence continues to support the recommendations.

Decision making for conditions being evaluated

A primary consideration in evaluating conditions is the availability of the test. The parameters that determine "availability" are numerous and vary considerably among conditions. It is also difficult to compare tests because of the differing "value" of a technology (e.g., multiplex capability, appropriateness of the site to conduct the screening service). The expert group considered whether the tests are amenable to a screening laboratory; for example, some tests are functional, such as those for hearing screening, and must be performed in the nursery. Other tests may have significant time constraints and are therefore better conducted in the hospital or birthing facility laboratory, as would likely be the case for bilirubin screening for kernicterus risk. It also should be noted that some of the conditions considered by the expert group did not meet the criterion that the test must be performed in the 24- to 48-hour period after birth (e.g., Wilson disease, familial hypercholes-

terolemia, Duchenne muscular dystrophy, congenital disorders of glycosylation, Turner syndrome screened by FSH levels). However, such conditions may be appropriate for screening at a later time in infancy or later in childhood. Although early and continuous screening of infants and children is a critical public health goal—as is lifelong screening—the expert group analysis was limited to conditions that should be and could be evaluated some time within the first few days of life. For the most part in the United States, the focus of traditional newborn screening programs has been on disorders detectable in the first 12 to 48 hours prior to discharge from the nursery. As such, the analyses were all predicated on testing done during this time frame. Initial screens in the neonatal period (i.e., first 28 days of life) would constitute a separate program with different costs and yields of cases and therefore should be separately analyzed.

Within this framework, the basis for decision-making as shown in Figure 9 starts with whether a screening test is available, a criterion without which decisions to screen cannot be made. Clearly, the first decision to screen is based on the availability of a sensitive and specific screening test that can be done in the 24- to 48-hour interval after birth. However, there is occasional disagreement as to whether a test is adequately validated for use in general populations. Hence, survey respondents may not necessarily give a 200-point score but may give a score between zero and 200. We defined the existence of the screening test as corresponding to a score between 100–200 points. Conditions determined to have a screening test are then evaluated with respect to the criteria.

Understanding that the evidence for each criterion needs to be evaluated, conditions with validated scores, scoring above 1,200 are considered appropriate for inclusion as primary targets in a screening program. However, the expert group distinguishes between those that are primary target conditions and those that are included in the differential diagnoses for those primary target conditions. Those with tests available and scoring between 1,000 and 1,200 are secondarily reconsidered as to whether an efficacious treatment is available and, if so, they are then reconsidered as to whether the natural history of the condition is well understood. If one of these is answered “no” but the condition is part of the differential diagnosis of a core condition, it is placed in the secondary target category. If it is not part of the differential of another core panel condition, the condition would not be considered appropriate for newborn screening at this time. Conditions falling between 1,000 and 1,200 are also considered appropriate for the secondary target category while those with an overall score under 1,000 are not considered appropriate for newborn screening at this time. At the bottom of the algorithm, the expert group acknowledges that there are currently significant research studies and clinical trials in process involving screening tests and therapeutics for diseases that might make the condition amenable to newborn screening (e.g., lysosomal disorders). The information that determined the current recommendation of the expert group is not static. Conditions not considered appropriate for the core

panel at this time should be reevaluated periodically to determine if their status has changed.

The data collection instrument used in this project provides information on only one aspect of a broader decision-making process required for evaluating conditions and establishing a uniform newborn screening panel (see decision tree in Fig. 9). There are also features of tests, such as costs, that are not factored into this diagram that State newborn screening programs may take into account. The algorithm can be used prospectively as a tool to evaluate conditions for their appropriateness for addition to or removal from a screening panel (Appendix 2). Reference information about each condition the expert group evaluated and the summary information can be compared to the results of an independent assessment of a condition. Review of the scientific literature should be conducted and expert opinion should be gathered for any condition evaluated. The preference is to use data from the literature. For the most subjective criteria, expert opinion is supplemented with the views of individuals involved with newborn screening programs and child health professionals and families.

Reporting responsibilities

Many factors affect the decisions about reporting of individual test results made by laboratories and programs. Some State newborn screening programs report directly to child health professionals, while others report to designated subspecialists. Some also report test results to families. Reporting also varies according to whether the results are screen-positive or screen-negative. As noted earlier, all results of likely clinical significance that are apparent in the testing platforms targeting specific conditions should be reported. As recommended by the Sickle Cell, Thalassemia and Other Hemoglobin Variants Subcommittee of CORN (1995), each screening program should develop guidelines for follow-up of carriers of all clinically significant conditions. This currently includes hemoglobinopathies and also would now apply to CF, because for both conditions the primary- or second-tier tests reveal carrier status. Similarly, second-tier testing for molecular causes of MCAD and other disorders can lead to the identification of carriers of the conditions (for autosomal recessive disorders). The differences in expectations between the conditions in the core panel and those in the secondary target category should be noted. Inherent to conditions in the core panel is the need to maximize detection in screening while minimizing excessive false positives being referred into the health care system. For conditions in the core panel that are positive on screening due to specific analytes being elevated, the secondary targets are identified in the diagnostic laboratory. It was on the basis of firm knowledge about these conditions that most decisions were made. The identification of conditions in the secondary target category is based on the fact that results are available due to the multiplex or multianalyte nature of the screening technology used. However, it does not presume that screening tests have been maximized for the detection of these conditions or that the knowledge base is sufficient to have developed an expectation of maximum health outcomes following interventions.

Newborn screening program officials also make decisions about following patients after initial screening and reporting. For instance, false-positives are treated as true positives until proven otherwise. However, once shown to be a real false-positive result, the State newborn screening program often treats the infant as they would a screen-negative infant, without pursuing further follow-up. The expert group believes that this situation warrants additional postconfirmation decision-making but acknowledges that the programs must minimally understand final diagnoses in order to discriminate false-positives from real-positives for these “secondary” targets.

State programs must decide whether the individual prevalence, costs and burdens of identifying these additional diseases—which may not be treatable and may take resources away from the treatable diseases originally targeted through these programs—can justify their inclusion in the program. They also must take into consideration the issues raised by child health professionals who will receive results about very rare conditions about which they have limited knowledge. Regardless of whether the State newborn screening program chooses to integrate secondary target cases into their full newborn screening program, it is important that an organized system of data collection and surveillance be available. The issues in newborn screening are similar to those that the FDA has faced with therapeutics for rare diseases, in which a shift toward phase IV (postmarket) surveillance during clinical trials has emerged. This shift recognizes that the most critical data about genetic diseases arise in the context of full population analysis. However, clinical data about the “normal” population is very scarce because the research focus has been on those with disease and on the diseases themselves. The significant variability inherent in genetic diseases requires significant knowledge of the expression of genetic variants in a general population before they are well understood. Such data collection has not been a priority of funding agencies.

E. Summary

Significant variability exists in the types of newborn screening available and the conditions screened across the United States. This project was intended to evaluate the scientific and medical evidence in order to identify conditions appropriate for newborn screening. After articulating overarching principles to guide decision-making, the current practices and systems in the States/regions and other countries were assessed.

All analyses were done from the perspective of national data, since one of the goals of the project was to bring standardization and uniformity to newborn screening. It is appreciated that some conditions may occur more commonly in subpopulations, such as is the case for IBG and HbE in the Hmong population, and that that may alter local screening practices.

Criteria were defined that would be used to compare the many conditions under consideration. The scientific literature related to each criterion was reviewed for each of 84 conditions and the opinions of at least three acknowledged experts for every condition was evaluated. At the first level of analysis, an assessment was made as to the availability of a screening test

that had been validated in a large general population. Scores were then established for each condition and they were assigned to one of three groups:

1. Core Panel (shared in common a high score [$\geq 1,200$], the availability of an efficacious treatment, a knowledge of natural history adequate for inclusion in a public health screening program);
2. Secondary Targets ([1,000–1,200] conditions that are part of the differential diagnosis of a core panel condition); and
3. Not Appropriate for Newborn Screening ($< 1,000$] either no newborn screening test is available or there is poor performance with regard to multiple other evaluation criteria).

The scientific evidence was overlaid on an initial categorization of conditions to ensure that all conditions in the core panel had a sufficiently well understood natural history and that an efficacious treatment was available.

The expert group recommends that State newborn screening programs:

1. Mandate screening for all core panel conditions defined by this report;
2. Mandate reporting of all secondary target conditions defined by this report and of any abnormal results that may be associated with clinically significant conditions. Some are identified in screening laboratories (e.g., hemoglobinopathies) and others in the diagnostic laboratory (e.g., MS/MS screened conditions). Clinically significant conditions also include the definitive identification of carrier status;
3. Maximize the use of multiplex technologies; and
4. Consider that the range of benefits realized by newborn screening includes treatments that go beyond an infant’s mortality and morbidity.

SECTION II: THE NEWBORN SCREENING SYSTEM: PROGRAM EVALUATION, COST-EFFECTIVENESS, INFORMATION NEEDS, AND FUTURE NEEDS

A. The newborn screening system

In order to successfully expand the number of mandated disorders screened for in newborns, the full breadth of the screening process and its components must be fully operational. Thus the expert group and its Diagnosis and Follow-up Work Group sought to examine the current status of screening systems throughout the United States, with particular attention paid to the diagnosis and follow-up components and their interface with the newborn screening program and primary health care professionals. In addition, the group was interested in identifying the key components of screening and highlighting some best practices that appear to improve outcomes. The six components of the newborn screening process that were assessed are:

1. Education, including prenatal education;

2. Screening, including specimen collection and testing;
3. Follow-up, including result reporting;
4. Diagnostic confirmation;
5. Management; and
6. Program evaluation and continuous quality improvement.

Much of the information reported in this section was obtained from a survey of State newborn screening programs conducted by the NNSGRC and reported at a November 2002 meeting sponsored by HRSA/MCHB and University of California, Los Angeles (UCLA), entitled "Educating Parents and the Informed Decision-Making Process Regarding Newborn Screening Procedures and the Use and Storage of Residual Bloodspots." NNSGRC has updated this information through June 2004.

Education

As screening increases there is a growing need for education across all groups of constituents, including parents and guardians, obstetrical providers, infants' medical homes, pediatric specialists, and emergency room/labor-delivery/neonatal intensive care unit (NICU) staffs. Education should occur in several places and times in the screening system, appropriate to the needs of patients, families, and health professionals.

Newborn screening programs typically provide educational materials during the perinatal period. The materials include information about newborn screening in general and brief descriptions of the conditions that are screened. Nineteen of 50 programs indicated that distribution of their newborn screening brochures was mandatory in birthing hospitals. Only one program reported not having an informational newborn screening brochure. All but three of the 50 programs indicated that their brochures included a list of disorders screened, and all but two described the specimen collection procedures and timing. Twenty provided information about when results would be available, 31 discussed the manner in which the results were reported to physicians, and 36 indicated how parents might obtain these results. As the number of conditions included in screening continues to expand, there has been a move toward providing more general information about the types of conditions screened rather than detailed information about each condition.

Prenatal Education

Few programs actively support education programs about newborn screening during the prenatal period. Ten of 50 State programs reported that newborn screening brochures typically were distributed in obstetrical offices, and 14 of 50 indicated that there was routine distribution in birthing classes. No information was available concerning quality, readability or understanding of the brochure information. The growing number of conditions for which newborn screening can be expected, combined with the existing limitations (e.g., familiarity of child health professionals with the newborn screening system) to delivering education during the perinatal period, argues for a focus on enhanced education during the prenatal

period. This area of need is currently being addressed by HRSA/MCHB through a contract with UCLA.

Screening

The timing of specimen collection and delivery to laboratories also varied. According to the NNSGRC 2000 National Newborn Screening Information Report, which included information from 28 programs at the time of this report, 74% of newborns were known to have been screened prior to 48 hours of age and 22% were screened after 48 hours. Twenty-two States reported that 2.7% of infants were screened prior to 12 hours of age, and 12.2% were screened between 12 to 24 hours of age. In several States as many as 30% to 40% of infants were screened between 12 and 24 hours of age. These timing issues may have direct implications for the predictive values of testing for some conditions.

Information about the timing of specimen delivery to laboratories was not readily available. The majority of programs rely on the United States Postal Service for specimen transport, with service varying from overnight delivery to up to a week in some areas. Most specimens arrive in the laboratories within 72 hours. However, in United States territories, such as Guam and States with relatively isolated and rural populations, delivery may take a week or more. It is suggested that specimens be transported by courier services that allow for receipt at the testing laboratories within 24 hours.

The timing of specimen collection and delivery is variably tracked. For diagnosed cases, programs generally record date of birth, date and time of specimen collection, date of receipt in the screening laboratory, date of laboratory report, and date of diagnosis. However, since establishing an etiologic diagnosis may be an iterative process that increasingly refines diagnosis, it can be difficult to define the time at which "diagnosis" is established. The date when initial diagnostic tests are ordered has been used as a substitute for date of diagnosis. Some programs monitor the date of initiation of treatment, but variations in the treatments for different conditions and the tendency to institute low-risk treatments in ambiguous, nonclassical cases renders this less useful unless viewed in the context of individual diagnoses. Most newborn screening programs presently operate on a 5-day work week. Some conditions can be life-threatening (e.g., MSUD, CAH, GALT, organic acidurias, fatty acid oxidation disorders, urea cycle disorders) within a few days after birth, so it is desirable to initiate specimen processing within 24 hours of specimen receipt in the laboratory, with a 5-day turnaround time between birth and the availability of the test results. However, it should be emphasized that detection of disease in the presymptomatic phase is one of the basic principles and values of screening.

The handling of screen-positive cases also was evaluated. Essentially, all newborn screening laboratories utilize a follow-up coordinator for reporting and tracking screen-positive results. For the most part, a positive result is reported only after the laboratory has verified the original finding through a second analysis of the original specimen. However, for some of the most time-sensitive conditions characterized by short-

term mortality and morbidity risks (e.g., CAH, galactosemia, isovaleric acidemia, MCAD, maple syrup disease, and some of the other metabolic diseases), preliminary positive results may be reported prior to repeat testing. These results are generally reported by telephone to the health professional identified by the newborn screening submittal form or by the birthing facility and/or the newborn screening consultant. The expert group recommends standardization of reporting procedures, including: the result, the reference range, the nature of the abnormality, and an indication of the speed and progression of clinical symptoms in the absence of intervention.

Screen-negative cases are often handled quite differently from the screen-positive cases. Some programs group normal results for batch reporting, waiting until all assays have been completed. Among the more significant potential problems identified in reporting of results is the risk of interpreting screening results as equivalent to diagnostic testing results. Screening results that are in the normal range may not have the same negative predictive value as is the case for diagnostic specimens obtained due to symptoms.³⁹ Additionally, it is increasingly apparent that age (developmental, chronological) and condition (acute affected, feeding status, transfusion status) of the newborn when the specimen was collected can affect the test results and their interpretation.⁴⁰

Further, the use of general terms such as “amino acids normal” or “acylcarnitines normal” in reporting of screen-negative results is an issue. The general lack of knowledge among clinicians of newborn screening programs and the screened conditions makes these types of results not useful. On the other hand, clinicians may not want to take the time to read through long, detailed, normal reports. A report indicating all that was normal in an MS/MS screening profile could require considerable information to reflect the varying degree to which different conditions had been ruled out. At the same time, it can be argued that detailed reports are necessary. For example, if an infant moves from one State to another that has a different screening panel, the results may be misinterpreted if they refer to a general group of tests rather than being delineated by condition.

The fact that two categories of screening tests and result reporting are proposed also complicates this issue. States vary in which primary-target conditions they choose to detect and the technology they use to detect them. In addition, there is variability in the testing strategies (e.g., use of second tier testing) and the cutoffs the program chooses to define cases. Diagnosis and Follow-up continues to consider these reporting issues.

Most programs report screened-negative results to the location identified on the newborn screening collection card, which in many cases is the hospital of birth and not necessarily the infant’s medical home. It has been observed in NNSGRC reviews of newborn screening programs that many hospitals do not routinely track the results and when the test results arrive at the hospitals, they are simply filed in the medical records without review. In addition, the tracking of newborn screening results to ensure that results are obtained on all

screened newborns, while desirable, is not a uniform hospital practice. As screening expands for the pediatric population, the medical home should consider incorporating verification status of newborn screening results and keep such records easily accessible in a manner similar to those used for posting immunization status to medical records. Recent efforts by HRSA/MCHB to support the development of integrated and linked information systems that include newborn screening information for health care providers’ direct access is an important development that may improve communication of screening results to the medical home and other appropriate health care facilities for the newborn. Additionally, national standards for the reporting of newborn screening results should be considered (similar to ACMG guidelines for prenatal DNA and other test report guidelines).

The use of second- or third-tier testing also was addressed in the work group’s assessments. This practice is fairly common in newborn screening laboratories. Almost all States use a second-tier test for CH, either T4 or TSH depending on which was used in the initial screen. These second-tier tests are commonly done on the original bloodspot sample and are distinguished from repeat testing, which involves repeating the same test on the original specimen, or second tests that require a fresh sample. Some programs use a second-tier fluorometric test following an initial bacterial inhibition assay for PKU. DNA testing as a second-tier test to detect high-frequency mutations is done in some programs for CF, hemoglobinopathies, MCAD, LCHAD and galactosemia, and some are considering second-tier testing by MS/MS for CAH. With expanded newborn screening (including hearing loss screening) identifying as many as 1:250 newborns who will require diagnostic confirmation (B. Threll, personal communication), the need to assess the capacity of the follow-up system is apparent.

Procedures for repeat testing in the newborn screening laboratory on the original bloodspot also were assessed. Essentially all newborn screening testing laboratories employ a QA step of retesting the original spot to confirm preliminary positive results. Some laboratories use a different method on second tests as a QA check. Retesting original bloodspots is distinguished from second-tier testing using a different test, and also from repeat screening, which uses a new specimen on which confirmatory testing is done. Routine repeat screening of all newborns is required in eight States, and several others strongly suggest second screening. There are specific circumstances (e.g., unsatisfactory specimens, acutely ill newborns in the NICU) under which repeat screening is commonly required. Because of the possibility of biologic false-positives, 29 States recommend/require a second specimen if tested prior to 24 hours of age and seven States require a second specimen if the newborn is tested before 48 hours of age. False-positives for CH and CAH are common in premature infants but can be dealt with through retesting when the infants are a few days older and their endocrine systems are more mature. Improved testing specificity on the initial specimen also can be achieved by using a nomogram more specific to the gestational age of the infant. False-negatives are the greater concern, since they may

not be recognized easily. Programs that mandate a second test for CH report finding 5% to 15% of their total caseload through the second test, but these cases have not been studied. This number is reduced by about 50% when TSH is used as the initial screening analyte. Over half of the cases of the classical simple virilizing form of CAH may go undetected on an initial screen due to biological factors.

Reporting and Follow-up

Follow-up is the term commonly used to describe the process of reporting abnormal screening results to the medical home, specialist, and/or guardians/parents and the initiation and tracking of the next steps in evaluation. Follow-up can be divided into two categories, short- and long-term follow-up. Short-term follow-up includes those activities that ensure all infants are screened, abnormal results are appropriately and expediently handled, and affected infants are promptly identified, appropriately referred, and treatment initiated where applicable. Long-term follow-up extends the period of follow-up substantially to monitor continuously the medical management and care coordination of those affected who require such services. Long-term follow-up also allows assessment of efficacy, sustainability, and safety of early treatment intervention, and can uncover new disease/treatment outcomes, and is valuable for demonstrating utility or limitations of screening.

Newborn dried bloodspot screening follow-up generally has functioned independently of newborn hearing screening follow-up, although many aspects of the follow-up procedures are similar and sometimes duplicative in terms of effort. Programs should minimize the number of places to which health care professionals must go to get information about their patients. Advances in information technology would allow direct and immediate access to screening test results, benefiting infants, health care professionals and screening programs. The experience of the newborn dried bloodspot programs could inform the hearing screening programs that have significant loss to follow-up of patients.

There is also some variation in how programs follow-up unsatisfactory specimens. Some State laws and program regulations place the responsibility for a satisfactory specimen on the specimen submitter. In such cases, the program tends not to pursue unsatisfactory specimens, electing to let the submitter perform its responsibility to the program. It is not clear that such practices had any impact on the liability issues that seem to have been the reason for such program practices to have arisen. In other cases, programs exercise their follow-up responsibilities in much the same way as they handle screen-positive cases. CLIA regulations require that a testing laboratory show that it has a procedure for improving specimen submissions in instances where there is unsatisfactory performance on the part of the specimen submitter.

Inadequate demographic information (e.g., patient's name, weight or age at the time of collection) also may render a specimen unsatisfactory. Most programs lack a strict enforcement policy regarding specimen rejection related to their rules governing certain demographic information. Often the initial re-

sponsibility for determining the acceptability of the specimen's demographic information falls to the clerical personnel performing the check-in process.

In order to improve the overall quality of specimens provided to newborn screening laboratories, the best approach is to minimize the number of unsatisfactory specimens and to ensure that an appropriate submitter education program is in place. It is best to have a designated person responsible for monitoring the quality of infant demographic information and for ensuring that accurate and complete information is part of a total quality management approach to laboratory operations. Compliance with requests for specimen demographic information must be monitored and action must be taken regarding noncompliance.

Most large States use computerized follow-up systems. Because these systems can be adapted to automated error surveillance, programs are encouraged to pursue routine quality checks using their computer systems. In the few States with computer generated submitter profiles, the profiles are used to improve the quality of specimens and information submission by, for example, monitoring periodic error rate reports. Those using computerized reporting and tracking systems have reported improvements on the part of submitters when profiling reports are used and submitters receive feedback from the reports.

In the event of a screen-positive result, most programs rely on information submitted with the newborn screening specimen to identify the newborn's physician or medical home. However, many newborns lack an identified child health professional at the time of release from the hospital. Often, the demographic information submitted with the specimen lists the nursery physician or on-call physician as the physician of record. Although identifying the appropriate child health professional may be a challenge, most newborn screening programs attempt to meet this challenge. Contact with the subspecialists is usually easier, since the group is smaller and is usually more intimately involved with the newborn screening program. In the interest of further closing the gaps in the system, it would be useful if hospitals were able to ensure that a follow-up appointment has been made for all newborns prior to their hospital discharge. At a minimum, the hospital nursery staff should work with families to identify the infants' medical homes and ensure that contact information for all infants is up to date.

Once the screen-positive case has been referred into the health care system, most programs have follow-up protocols that include tracking the patient until treatment has been initiated. Some programs subcontract this responsibility to regional medical centers and do not actively pursue this information, having transferred the responsibility for this in their contracts. However, this practice may complicate ready access to short- and long-term information that would be useful for program evaluation. Some States are developing systems that allow information integration and program linkage to improve tracking of screening results and patient outcomes. For example, some use bar codes that link newborn screening filter paper cards with birth certificates, and others have considered

including the newborn screening information on the face page of the medical record where vaccination information is placed to facilitate monitoring. In any case, a plan should be in place for exhaustive and documented confirmation of follow-up. Follow-up coordinators should link repeat specimens to initial specimen records, and all programs should obtain short- and long-term follow-up information.

A variety of methods of screen-positive results notification have evolved within newborn screening. In most programs, once the follow-up coordinator has provided results to the child health professional, the child health professional or a member of his or her staff informs the family of the screening results. Some programs notify both the child health professional and the family. Education is an important aspect of the notification of parents and health care professionals. Some States have developed culturally and linguistically appropriate educational materials for families but there is limited availability of similar materials for child health professionals and specialists.

Once the family is informed of the test results, the child health professional determines the need for and extent of subspecialty involvement, unless the program's follow-up is conducted directly through subspecialists. Not all conditions have similar demands for the timeliness or complexity of follow-up. The availability of informational materials for child health professionals that would facilitate their ability to participate actively in a collaborative management approach to their patients' care would be useful. Such information could include immediate management issues and relevant subspecialist referral sites. The work group on Diagnosis and Follow-up developed templates for such informational materials that have been pilot tested at limited sites. They are the basis of ongoing work developing templates for all conditions in the core panels, as well as those in the secondary target category. (Examples of these templates can be found in Appendix 3.) Although guidelines for immediate management could be readily developed, there is little standardization of parameters by which one would qualify an experienced subspecialty provider. Further, some parts of the country may have limited availability of experienced pediatric and subspecialty care health care professionals. This is particularly apparent in the area of inborn errors of metabolism; there are currently 53% fewer board certified biochemical geneticists in the United States than were practicing in 1990 and a limited number of trainees. In such circumstances, an organized system to link child health professionals with specialty care professionals would be useful. This could be accomplished through the developing HRSA/MCHB Genetics and Newborn Screening Regional Collaboratives that are intended to make national and regional services and resources accessible at the local community level.

Once confirmation of diagnosis is available to the child health professional or subspecialist, it is common for this information to be communicated promptly to the State newborn screening program. It is important that all programs obtain confirmatory outcome reports in order to fulfill their public health mandate.

Diagnosis

There is a complex relationship between the definition of screen-positive test results and the definition of the genetic condition itself. Upon identifying a screen-positive infant, algorithms through which diagnostic confirmation is obtained are followed. Some steps may involve the screening laboratory as is the case with second-tier tests while others involve the clinical and laboratory evaluations that lead to the final diagnosis. It is only after significant testing in a general population that the full breadth of the phenotype of the genetic condition in question is well understood. Hence, it becomes important to maintain communication between the health care professionals and the screening programs related to the false-positive and true-positive results. It will also be important to reconsider what constitutes a false positive result since a particular screening result may be associated with either a core condition panel or a secondary target condition. Further, it is important to develop mechanisms through which programs can be made aware of patients identified outside of the program in order to adjust program parameters to avoid "missed" cases. Finally, given that genetic tests can provide information about affected individuals and carriers, clear policies should be in place about communicating such information.

Management

Many programs do not have educational materials to facilitate and optimize patient care once a patient is diagnosed. Such information is commonly in the purview of the experts who develop guidelines for treatment. Information dissemination practices that facilitate collaborative management between the child health professionals and specialists would be useful.

Over the longer term of intervention and treatment there is usually insufficient information shared between health care professionals and the programs, and contact beyond the initial treatment phase is rare. This gap might only be filled through the development of information collection systems that facilitate the integration of program information with other health care information.

The availability of and access to therapeutic interventions varies among the States. Some States provide funding for medical foods[†] either completely or on a sliding scale based on income. Costs not covered by insurance may be covered through Title V funds and Medicaid. However, they are less likely to fund genetic counseling, penicillin for sickle cell disease, or thyroid hormone replacement therapy.

A definition of the range of health care professionals considered necessary for managing a particular condition is limited. Medical and nonmedical services are generally defined by the health care professionals to whom the infants have been referred. However, because almost all programs provide no funding for health outcome evaluation, few long-term studies exist. Beyond one to three years of age, there is little coordinated or systematic monitoring by the programs.

Program Management

Programs use a mix of models for management and development of their newborn screening activities. Many States have external advisory committees, although some rely only on internal advisory groups, which may not include consumers and experts for conditions considered by the programs.

B. Program evaluation

Several of the goals of this project are aimed at standardizing language and identifying the data or information needed to evaluate newborn screening program performance. Historically, newborn screening programs have been evaluated only internally, with the exception of the screening laboratory, which generally must meet CLIA requirements even though some of the analytes may not be specifically covered. Since 1987, HRSA/MCHB has made available to the States consultative program reviews by a team composed of experts in various aspects of newborn screening activities, and this has been continued as a responsibility of the NNSGRC. Besides providing annual State data specific to the Title V Block Grant performance measure, programs voluntarily report their program performance data to the NNSGRC for compilation and publication as an annual newborn screening data report. These reports are available at the NNSGRC website and can be used for inter- and intraprogram comparison (See www.genes-r-us.uthscsa.edu). Uniform performance measures, however, could enable better and more standardized comparative assessment of newborn screening programs. Performance standards should be related to the needs of those with the specific conditions identified. Uniformity of language and standardization of performance measures will allow programs to move from independent evaluation to a comparative system targeted at high quality and efficiency.

Program Standards

A fundamental goal of newborn screening is benefit to the newborn by identifying a treatable condition. Variability exists among the conditions in the core panel regarding the speed with which they must be treated in order to minimize or eliminate the negative consequences of the condition. In newborn screening programs, speed of screening and reporting results is sometimes driven by the conditions that have the most demanding time needs. For example, an elevated 17-hydroxyprogesterone indicates a high likelihood that classical CAH is present and should therefore be pursued promptly, since in some instances death can occur from salt wasting within the first two weeks of life. Similarly, an elevated C8 acylcarnitine indicates a high likelihood that MCAD is present and should therefore be pursued promptly, since in some instances death can occur within the first two weeks of life. This contrasts with the finding of hearing loss, for which the interventions can be delayed for two to three months without significantly affecting speech development. The importance of education of families and the medical home about timing and the consequences of later notifications is apparent.

Appendix 4 lists specific steps in the newborn screening program process that should be monitored. Program performance can be improved by integrating data monitoring into policies and procedures and then modifying programs as problems are identified. Furthermore, development of a uniform approach to data collection and program evaluation allows for the comparison of program performance among States.

National Programs of QA

On a national basis, there is no comprehensive QA program for newborn screening aside from that provided for screening laboratories by CDC (see Fig. 10). CDC offers a proficiency testing and quality assurance program specifically for newborn screening laboratories—the Newborn Screening Quality Assurance Program. The newborn screening laboratories are regulated under CLIA of 1988. FDA provides additional oversight of manufacturers who provide testing products to newborn screening laboratories, and CDC provides a service that validates the filter paper bloodspot collection devices. The NNSGRC, funded by HRSA/MCHB, provides consultative program reviews that include all aspects of the newborn screening system (upon the official invitation of individual State newborn screening programs), and collects and assimilates national newborn screening data.

The Joint Commission on Accreditation of Hospital Organizations (JCAHO) plays a role in the oversight of activities within hospitals. For several reasons, JCAHO's activities have not been specifically directed toward the hospital's role in newborn screening. Even though birth hospitals collect the vast majority of screening specimens, record demographic information, and receive newborn screening test results, hospitals have not traditionally been held accountable to JCAHO for newborn screening activities per se. Historically, hospital responsibilities for tracking newborn screening testing results have been varied, particularly since the newborns are usually not in the hospital when the screening results are completed and returned. Most State screening regulations are silent on hospitals' responsibilities, though some include specific requirements, and hospitals and administrators can in some States be held liable if newborn screening practices are improperly performed. Oversight of newborn screening has been complicated by the fact that the oversight of clinical activities is limited compared to the regulation of laboratories, which includes maintaining records of specimen submission and result reporting. In many hospitals, newborn screening specimens are collected and submitted to the screening laboratory directly from the newborn nursery, bypassing some areas of this laboratory oversight. Hospitals appear to assume greater responsibility for screening conducted within the nursery, for example, screening for hearing loss. In such circumstances, hospitals have a clear responsibility to make patients aware of any critical laboratory information stemming from their hospital stay. However, since hearing screening results are immediately available, the task of initiating notification and arranging for next steps in evaluation is simplified.

Discussions are ongoing regarding the possibilities of improving the ways in which hospitals provide information to newborn screening programs to ensure that adequate information is available in a timely manner for recontacting families or health care professionals and establishing follow-up while still maintaining appropriate privacy of the patient's medical information.² At the level of diagnosis and follow-up, there are several programs that have worked toward ensuring quality. Some organizations, such as CORN, AAP, ACMG, and the Society for Inherited Metabolic Disorders (SIMD), have been involved in the development of practice guidelines for the diagnosis, treatment, and management of many of these conditions. In addition, there are programs with "deemed" status through CLIA that offer proficiency testing and inspections of the laboratories providing diagnostic services for the conditions included in newborn screening programs. However, at the present time most analytes that are screened are not included in this program, although their addition is under active discussion.

Some programs have developed internal QA programs that variably address the components of the newborn screening system. While all States tabulate the number of tests done, many cannot relate tests to birthing records in order to ascertain the percentage of newborns screened. On the other hand, programs routinely track time from birth to diagnosis and treatment, and the numbers of newborns lost to follow-up, which are extremely important aspects of the screening system. Most programs maintain records of unsatisfactory specimens but they vary in follow-up actions and educational programs to improve specimen quality. In this respect there is perhaps a role for the federal government in providing some form of national program oversight. Furthermore, there are very different forms of oversight for laboratory services than for clinical services. In order to continue to improve the quality of newborn screening programs, several actions should be taken:

1. There should be uniformity in the types of data collected (see Appendix 4) by programs in order to compare program performance among States. In addition, reporting to a central authority should be required.
2. Periodic performance reviews of all components of newborn screening programs should be required. This should be a federal responsibility.
3. Language and terminology should be standardized in order to better compare performance among programs.
4. Turnaround time in reporting screen-negative results should be improved.
 - a. At a minimum, all results from the initial screening test (some States perform a second test later) should be available less than five days after the blood sampling for the first posthospital discharge visit to be of use in this clinical visit and to facilitate awareness of lifelong screening. Most results should be available within two days of the specimen arriving in the laboratory, and specimens should arrive in the laboratories within three days of collection.
5. Diagnostic laboratory QA programs should be enhanced to include all conditions screened in newborns.
6. Organized systems to allow for the collection and analysis of data about patients are important in defining the standards to be met and improving our understanding of these typically very rare conditions. Data from population-based screening are the optimal source of unbiased information about conditions and required reporting should be instituted.
7. Hospitals and JCAHO have significant roles to play, and standards need to be developed to improve quality, minimize errors, and facilitate tracking of newborns requiring active participation in testing follow-up.
8. All newborn screening laboratories should be CLIA-certified and should participate in CDC and CAP/ACMG proficiency testing programs or other equivalent programs as applicable.
9. All States should have an active system-wide newborn screening QA and total quality management program.
10. To bring uniformity to programs across the country and thereby create a more equitable system for all Americans, national oversight and authority must be provided with adequate resources. Consideration should be given to institutionalizing the role of the HRSA-funded NNSGRC, which currently offers on-site expert consultative reviews to the State newborn screening programs.

C. Cost-effectiveness analysis

This project focused primarily on a scientific analysis of conditions and the features that should be considered when deciding whether they should be included in a newborn screening program. However, costs often are the basis on which such decisions are made. Review of the few available cost-effectiveness studies of newborn screening suggests that often, they may be too limited in scope. Some studies have focused on the short-term costs and benefits of the screening stage and the immediate steps following the identification of a screen-positive infant. Most address tests for only a small number of disorders, and none has explored the cost savings and clinical benefits of tests such as MS/MS.⁴¹⁻⁴⁶

A basic cost-effectiveness analysis was conducted to better inform our decisions. Costs and benefits related to screening for particular conditions or groups of conditions were evaluated after mapping them over major disease outcomes (e.g., life expectancy, cerebral palsy/stroke, seizures, developmental delay, hearing loss, vision loss). Costs were obtained from the literature.^{2,42,43,47-51} Benefits were determined from expected outcomes with and without early treatment or intervention. Quality-adjusted-life years (QALYs) were then compared to costs. Where appropriate, tests capable of being multiplexed with other tests for different conditions were assessed independently and as a group. Results were found to be stable by sensitivity analysis.

The results of these analyses indicate that all newborn screening programs evaluated improved outcomes and most reduce overall costs (Carroll and Downs, in press). Screening

for CAH added increased cost per QALY gained, but the cost was well within the range conventionally considered cost effective. Screening for galactosemia was the only strategy that would be considered not cost effective in the base case analysis. However, under some reasonable assumptions, it can be shown to be cost effective. The identification of potentially affected individuals at such an early time in life leads to many years over which the benefits accrue and, in aggregate, the benefits outweigh the costs.

Technologies such as MS/MS further save money due to their multiplexing capability and low screening false-positive rates. MS/MS, used to screen for multiple conditions, had the greatest impact on outcomes and saved the greatest amount of money in the analysis. Virtually all screening for conditions that are treatable with significantly beneficial outcomes can be justified with benefits increasing as more conditions are included. The analysis also showed that clinical benefits and savings depend on low false positive rates and timely follow-up and treatment of positives, emphasizing the importance of an integrated screening and follow-up program.^{41–45,52}

D. Information gaps and a research agenda

Data and Analytical Needs

Screening

The evidence base for disorders potentially amenable to screening is limited and the questions that must be answered to inform our decisions about the future of our newborn screening programs are numerous and the issues complex. There are cutting edge new technologies emerging that can have a significant impact on screening programs. However, tech assessments have limited capacity to identify issues about promising technologies early in their development (e.g., is there sufficient capacity in the system to test the 4.1 million United States newborns? Are the tests adequately validated?). This raises important questions about how to implement new technologies for screening. Historically, as new technology is validated on a known cohort, it is then applied to a prospective screening cohort in a linked or unlinked (e.g., HIV screening) method, with or without reporting, and with or without randomization (e.g., CF). Many State newborn screening programs have awaited data from other State pilot or trial programs before investing in the costs of incorporating new technologies into testing and follow-up protocols. The potential for screening beyond the first few days of life is increasing. Determining how best to link existing public health activities (such as immunization) that occur at specific clinical points later in life offers opportunities to screen for additional conditions that are less amenable to screening in the first 24 to 48 hours of life. Information technology has opened up opportunities to improve the systems that support the medical home's integrated role in newborn screening and there is always the opportunity to improve informatics and communications and their integration into public health information systems and registries.

There is an ongoing and growing need to articulate a research agenda for the many conditions that are already part of

newborn screening. For example, the impact on the optimal timing of screening of newborns in the neonatal intensive care unit that have received hyperalimentation or packed cell transfusions remains unclear.

Follow-Up

Many questions remain about the impact of screening for a larger number of rare disorders, as well as what the true significance is of a “false-positive” or “transiently abnormal” screening test.⁵³ These may require costly, long-term evaluation projects in order to obtain the statistical power needed to better understand these issues in rare diseases. Again, we may need a broader national approach to data collection and analysis.

Diagnosis

Considerable research potential exists in the area of diagnosis of these rare diseases. The preferred approaches and methods of diagnosis and confirmation of presumptive diagnoses remain to be determined and our understanding of the natural history of the conditions and the associated genotype-phenotype correlations can only improve. There are many questions to be answered for each of the conditions for which screening is currently offered. For instance, there is still little information available about the outcomes of infants identified in G6PD screening programs. The interrelated roles of genetic risk factors and the environmental exposures that trigger disease expression are areas where large collaborative research projects will be needed. The use of the National Children's Study as a component of newborn screening research offers a number of opportunities. Similarly, we need to understand the issues and barriers that lead to the lack of hearing screening follow-up to determine etiology.

Management

The emerging area of collaborative disease management offers many opportunities to improve our newborn screening programs. The nature of our health care system is such that the bridges between child health professionals and specialists must be strengthened. Issues of interest include: 1) how best to partner with the medical home; 2) how to facilitate the transition to adult care (childhood cancer survivorship model); and 3) what are the expected outcomes for the adults with these now chronic diseases. It is also likely that situations similar to that of maternal PKU will arise with other metabolic diseases, such as 3-MCC, or the endocrinopathies, such as CH. Long-term outcomes research will require organized systems of data collection and monitoring. There are also gaps in our understanding of treatment issues for many conditions (e.g., nonclassical CAH). We also need to elucidate the long-term behavioral and educational issues associated with children with conditions detected by newborn screening.

Evaluation

Program evaluation can also benefit from organized collaborative research programs. The creation of registries for long-term outcomes research and for system validation offers a clear pathway to improvement of the programs.

Health Systems And Outcomes Research

Our health care system continues to evolve in parallel with the evolution of the newborn screening programs. The increas-

ing diversity of the United States population necessitates that health disparities research as relates to diagnosis, management, and long-term follow-up of patients identified in newborn screening be enhanced.

Education

The trend toward more direct consumer involvement in health care decisions and prevention indicates the need for enhanced educational programs for the public. Further, the rarity and complexity of the many conditions already screened suggests a need for improved educational programs for the professionals. Opportunities remain to improve our understanding of the primary communication and education needs related to a screen-positive result in newborn screening. Similarly, many questions remain about the issue of appropriate decision-making relative to newborn screening. There is a need to understand the issues that arise in the delivery of prenatal education and determine the best models for such education while still working to broaden overall genetics public education. There is also a need to improve our understanding of how attention to cultural diversity and literacy could contribute to effective newborn screening programs. In order to better understand the limitations of and impediments to education, best practices models related to who provides services (e.g., birth educators, obstetrician gynecologists, subspecialists) need to be identified and there is need to understand how they can be provided outside the delivery room or nursery, and when they are best provided. The role for cross-specialty education and continuing medical education for health care professionals is also an area that would benefit from study. Last, there is considerable opportunity for research into the ethical, legal, and social issues that arise with expanded newborn screening and newborn screening in general.

Health Systems As Related To Newborn Screening

A better understanding of the organization and functioning of our newborn screening related health care systems would also benefit the continued development of programs. In particular, studies of systems of care that would offer the highest quality delivery of newborn screening services would improve the programs.

Other

There are numerous ancillary issues that relate to improving newborn screening outcomes. These include: 1) expanding screening opportunities prenatally and after birth when timing of testing, identification, and intervention offer additional value for health outcomes in the pediatric population; 2) ongoing research efforts to identify better and new screening and intervention strategies for rare and common disorders; and 3) continued research into outcomes of transiently abnormal screens to determine if such test results have predictive value for later diseases as well as to measure the psychosocial impact of such results (e.g., costs of vulnerable child issues). Some of the diseases for which postnatal newborn screening is recommended may be additionally benefited by prenatal detection; however, prenatal screening is not presently universally available. We may gain a better understanding of the incidence and spectrum of diseases associated with perinatal and early child-

hood mortality by implementing uniform child autopsy policies and procedures which ensure availability of appropriate studies (including metabolic and genetic studies for all perinatal deaths, including stillbirths) and early unexpected childhood deaths.

E. Future needs

Hopefully all screening programs can benefit from a more robust national role and increased national standards and policies for newborn screening. Because so many of the conditions screened in newborns, or under consideration for screening, are rare, most States that undertake evaluations of the scientific basis for screening of conditions must rely on the same relatively small group of patients identified throughout the world. There is a potential national role in providing scientific evaluation of conditions and defining core condition panels. This would allow the States to apply the best science to their own considerations when determining their role in expanded screening. Practice guidelines also could be developed at a national level by interested organizations. There is also a potential expanded national role in oversight and enforcement, data collection, program evaluation, and the development of educational materials to support newborn screening.

Depending on the overall incidence of particular conditions, regional cooperatives should coordinate access to health care professionals, serve as coordinators and repositories for data collection, provide long-term follow-up capability when resources and expertise are limited, facilitate transition (and access) from pediatric to adult care, and provide education. The distribution of primary, secondary, and tertiary services is largely based on the incidence of a condition and the complexity of its short- and long-term diagnosis and management. For more common conditions with easier diagnosis and follow-up, there is likely to be sufficient local health care expertise for patient care. As incidence decreases and complexity increases—particularly for rare metabolic diseases—services become more difficult to access. Developing resources and infrastructure to ensure that health care professionals with appropriate expertise are available locally, regionally, and nationally will be important to ensuring access to high-quality services.

States also must retain their significant roles and responsibilities. They have a clear authority with regard to oversight and evaluation, as well as enforcement. There is a need to integrate the various systems of health care coverage and payment through flexible and comprehensive financing of services. Service coordination at both State and local levels must be considered, as well as program integration with the State Children's Health Insurance Plan, early intervention programs, Title V programs, Medicaid, and similar services.

In considering the national role in newborn screening, it is apparent that there are already significant barriers to the creation of a model newborn screening system in the United States. For example:

1. Financing across State and county lines is constrained by Medicaid rules;

2. Service delivery is fragmented on a disease basis;
3. There is lack of universal access and ability to access the medical home;
4. There is insufficient support to bridge geographic barriers;
5. It is difficult to identify experienced health care professionals for complex care (e.g., centers of excellence for genital reconstructive surgery for CAH; confirmation of metabolic diagnoses);
6. Misinterpretation of privacy regulations (e.g., HIPAA) (see Appendix 5 for discussion and clarification of HIPAA related issues in the context of a public health program);
7. There is underutilization and lack of uniformity of information technology;
8. Collaborative management and care is constrained by systems of reimbursement;
9. There is variability in State mandates;
10. State sovereignty sometimes dictates individual approaches; and
11. There is variability in financing of screening programs.

F. Summary

In order for expanded newborn screening to be implemented universally, a well operating and standardized newborn screening system must be in place. At the present time there is significant variability among the State programs with regard to policies and practices employed after screening and in initial notification of health care professionals. The expert group evaluated the components of the system and their associated functions with a primary focus on the parts of the system that interface specialty care professionals with either the newborn screening program or the child health professionals.

A basic cost effectiveness study of newborn screening was conducted. The results of this analysis demonstrated that newborn screening is cost effective when compared to other recommended medical expenditures. This supports the recommendations made in Section One of this report regarding the need to expand the breadth of conditions that should be included in core screening panels and the secondary target category.

The scientific analyses and systems evaluations also identified gaps in our knowledge base and pointed to areas in which research is needed. The expert group recommends that:

- Programs continue to improve the components of the system beyond the initial screening, communication of those results, and ensuring that the newborn enters into short-term follow-up. To accomplish this:
 - reporting procedures should be standardized
 - reports of confirmatory results should be obtained
 - There should be improved oversight (e.g., JCAHO) of the hospital-based screening activities to improve tracking of screen-positive cases;
 - There should be more uniformity in the language and definition of the performance standards (e.g., repeat test, second test) monitored and reported by programs;
- The QA programs involving the diagnostic and follow-up system should be enhanced;
- National oversight and authority with appropriate resources should be provided; and
- Systems should be in place for collection of data about individuals identified as screen-positive in newborn screening programs.

References

1. 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.
2. U.S. Congress, Office of Technology Assessment, Healthy Children: Investing in the Future, Appendix 1: Data and Methods Used in OTA'S Cost-Effectiveness Analysis of Strategies for Newborn Screening, 1988. Available at: <http://www.wws.princeton.edu/cgi-bin/byteserv.prl/ota/disk2/1988/8819/881919.PDF> Accessed April 7, 2006.
3. McCabe ERB. Principle of newborn screening for metabolic disease. *Perinatol Neonatol* 1982;6:63–73.
4. Therrell BL, Panny SR, Davidson A, Eckman J, et al. U.S. newborn screening system guidelines: statement of the Council of Regional Networks for Genetic Services (CORN). *Screening* 1992;1:135–147.
5. Therrell BL, Berenbaum SA, Manter-Kapanke V, Simmank J, Korman et al. Results of screening 1.9 million Texas newborns for 21-hydroxylase-deficient congenital adrenal hyperplasia. *Pediatrics* 1998;101:583–590.
6. Pass KA, Lane PA, Fernhoff PM, Hinton CF, et al. US newborn screening guidelines II: follow-up of children, diagnosis, management, and evaluation. Statement of the Council of Regional Networks for Genetic Services (CORN). *J Pediatr* 2000;137:S1–S46.
7. Dhondt JL, Farriaux JP, Saily JC, Lebrun T. Economic evaluation of cost-benefit ratio of neonatal screening procedure for phenylketonuria and hypothyroidism. *J Inher Metab Dis* 1991;14:633–639.
8. Tsevat J, Wong JB, Pauker SG, Steinberg MH. Neonatal screening for sickle cell disease: a cost-effectiveness analysis. *J Pediatr* 1991;118:546–554.
9. Wright L, Brown A, Davidson-Mundt A. Newborn screening: the miracle and the challenge. *J Pediatr Nursing* 1992;7:26–42.
10. Pollitt RJ, Green A, McCabe CJ, Booth A, et al. Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assess* 1997;1:1–202. Review.
11. Guthrie R. Screening for “inborn errors of metabolism” in the newborn infant – A Multiple test program. *Birth Defects* 1968;4:92–98.
12. Guthrie R, Susi A. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* 1963;32:338–343.
13. Academy of Pediatrics. Serving the family from birth to the medical home: A report from the Newborn Screening Task Force convened in Washington DC, May 10–11, 1999. *Pediatrics* 2000;106:383–427.
14. Wilson JM, Jungner YG. Principles and practice of screening for disease. Geneva: World Health Organization, 1968.
15. Levy HL. Newborn screening by tandem mass spectrometry: a new era. *Clin Chem* 1998;44:2401–2402.
16. Bailey DB. Newborn screening for fragile X syndrome. *Ment Retard Dev Disabil Res Rev* 2004;10:3–10.
17. Weatherall DJ, Clegg JB. The thalassemia syndromes 4th ed. Oxford: Blackwell Science; 2000.
18. Chace DH, Kalas TA, Naylor EW. Use of tandem mass spectrometry for multianalyte screening of dried blood specimens from newborns. *Clin Chem* 2003;49:1797–1817.
19. Meredith-Malloy N. Newborn Screening Quality Assurance Program: 2002 Tandem Mass Spectrometry Annual Summary Report. Atlanta: Centers for Disease Control and Prevention. 2003.
20. Rinaldo P, Tortorelli S, Matern M. Recent developments and new applications of tandem mass spectrometry in newborn screening. *Curr Opin Pediatr* 2004;16:427–433.
21. Marsden D, Larson CA. Emerging role for tandem mass spectrometry in detecting congenital adrenal hyperplasia. *Clin Chem* 2004;50:467–468.
22. Willi SM, Moshang T. Diagnostic dilemmas. Results of screening tests for congenital hypothyroidism. *Pediatr Clin North Am* 1991;38:555–566.

23. Lacey JM, Minutti CZ, Magera MJ, Tauscher AL, et al. Improved specificity of newborn screening for congenital adrenal hyperplasia by second-tier steroid profiling using tandem mass spectrometry. *Clin Chem* 2004;50:621–625.
24. Ranieri E, Lewis B, Morris CP, et al. Neonatal screening using combined biochemical and DNA-based techniques. In: Dodge JA, Brock DJH, Widdicomb JH, editors. *Cystic Fibrosis: Current Topics*. Chichester, UK: John Wiley & Sons, 1996:181–206.
25. Blau N, Thony B, Cotton RGH, Hyland K. Disorders of tetrahydrobiopterin and related biogenic amines. In: Scriver CR et al, editors. *The Metabolic and Molecular Bases of Inherited Disease*, 8 ed. New York: McGraw-Hill, 2001:1725–1776.
26. Report of a WHO Scientific Group. Screening for inborn errors of metabolism. World Health Organization Tech Report Series No. 401, 1968.
27. Holtzman A, Watson MS. Task Force on Genetic Testing: Promoting Safe and Effective Genetic Testing in the United States. Baltimore: Johns Hopkins Press, 1997.
28. Howse J, Katz M. The importance of newborn screening. *Pediatrics* 2000;106:595–
29. Frankenburg WK. Selection of diseases and tests in pediatric screening. *Pediatrics* 1974;54:612–616.
30. Therrell BL. U.S. newborn screening policy dilemmas for the twenty-first century. *Molec Genet Metab* 2001;74:64–74.
31. American Academy of Pediatrics. Medical Home Initiatives for Children with Special Needs Project Advisory Committee. The medical home. *Pediatrics* 2002;110:184–186.
32. American Academy of Pediatrics Steering Committee on Quality Improvement and Management. Classifying recommendations for clinical practice guidelines. *Pediatrics* 2004;114:874–877.
33. Genetic Evaluation of Hearing Loss Expert Panel. Genetics evaluation guidelines for the etiologic diagnosis of congenital hearing loss. *Genet Med* 2002;4:162–171.
34. Young SS, Kharrazi M, Pearl M, Cunningham G. Cystic fibrosis screening in newborns: results from existing programs. *Curr Opin in Pulm Med* 2001;7:427–433.
35. Merelle ME, Nagelkerke AF, Lees CM, Dezateux C. Newborn screening for cystic fibrosis. *Cochrane Database Syst Rev* 2001;3:CD001402. Review.
36. Council of Regional Networks for Genetic Services (CORN), Sickle Cell, Thalassemia and Other Hemoglobin Variants Subcommittee. Guidelines for the follow-up hemoglobinopathy carriers detected by newborn screening, 1995.
37. Refsum H, Fredriksen A, Meyer K, Ueland PM, Kase BF. Birth prevalence of homocystinuria. *J Pediatr* 2004;144:830–832.
38. American Academy of Pediatrics Subcommittee on Management. Clinical Practice Guidelines 2004. Management of hyperbilirubinemia in the newborn infant 35 or more weeks gestation. *Pediatrics* 2004;114:297–316.
39. Eaton RB, Comeau AM, Zytkevich TH, Larson C. Newborn screening: new developments in a proven field. *Clin Lab Sci* 2002;15:239–44.
40. Larson CA. Congenital Hypothyroidism. In: Pediatric Endocrinology: A Practical Clinical Guide. Radovick S, MacGillivray MH, eds., Totowa, NJ, Humana Press, 2003:275–290.
41. Panepinto JA, Magid D, Rewers MJ, Lane PA. Universal versus targeted screening of infants for sickle cell disease: a cost-effectiveness analysis. *J Pediatr* 2000;136:201–208.
42. Insinga RP, Laessig RH, Hoffman GL. Newborn screening with tandem mass spectrometry: examining its cost-effectiveness in the Wisconsin Newborn Screening Panel. *J Pediatr* 2002;141:524–531.
43. Keren R, Helfand M, Homer C, McPhillips H, Lieu TA. Projected cost-effectiveness of statewide universal newborn hearing screening. *Pediatrics* 2002;110:855–864.
44. Schoen EJ, Baker JC, Colby CJ, To TT. Cost-benefit analysis of universal tandem mass spectrometry for newborn screening. *Pediatrics* 2002;110:781–786.
45. Venditti LN, Venditti CP, Berry GT, Kaplan PB, et al. Newborn screening by tandem mass spectrometry for medium-chain Acyl-CoA dehydrogenase deficiency: a cost-effectiveness analysis. *Pediatrics* 2003;112:1005–1015.
46. 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;8:1–121.
47. PriceWaterhouseCoopers. Newborn screening programs: an overview of cost and financing, (January 2001) Available at: http://www.marchofdimes.com/files/Final_PWC_NBS_Report2.pdf. Accessed May 18, 2006.
48. Washington State Department of Health. Least burden and cost benefit analysis of newborn screening for metabolic disorders, (2003) Available at: <http://apps.leg.wa.gov/WAC/default.aspx?cite=246-650>. Accessed May 18, 2006.
49. Angus DC, Barnato AE, Linde-Zwirble WT, Weissfeld LA, et al. Robert Wood Johnson Foundation ICU End-of-Life Peer Group. Use of intensive care at the end of life in the United States: an epidemiologic study. *Crit Care Med* 2004;32:638–643.
50. Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment, United States, 2003. *Morb Mortal Wkly Rep* 2004;53 (3):57–59.
51. Carroll AE, Downs SM. A comprehensive cost-utility analysis of newborn screening. *Pediatrics* 2006;117:287–295.
52. Keren R, Helfand M, Homer C, McPhillips H, Lieu TA. Projected cost-effectiveness of statewide universal newborn hearing screening. *Pediatrics* 2002;110:855–864.
53. Fearing MK, Levy H. Expanded newborn screening using tandem mass spectrometry. *Adv Pediatr* 2003;50:81–111.