Newborn Screening: Toward a Uniform Screening Panel and System—Executive Summary

Michael S. Watson, PhD, Marie Y. Mann, MD, MPH, Michele A. Lloyd-Puryear, MD, PhD, Piero Rinaldo, MD, PhD, R. Rodney Howell, MD, American College of Medical Genetics Newborn Screening Expert Group

ABSTRACT

The Maternal and Child Health Bureau commissioned the American College of Medical Genetics to outline a process of standardization of outcomes and guidelines for state newborn screening programs and to define responsibilities for collecting and evaluating outcome data, including a recommended uniform panel of conditions to include in state newborn screening programs. The expert panel identified 29 conditions for which screening should be mandated. An additional 25 conditions were identified because they are part of the differential diagnosis of a condition in the core panel, they are clinically significant and revealed with screening technology but lack an efficacious treatment, or they represent incidental findings for which there is potential clinical significance. The process of identification is described, and recommendations are provided.

IN THE UNITED States, newborn screening is a highly visible and important state-based public health program that began >40 years ago. States and territories mandate newborn screening of all infants born within their jurisdiction for certain disorders that may not otherwise be detected before developmental disability or death occurs. Newborns with these disorders typically appear normal at birth. Appropriate compliance with the medical management prescribed can allow most affected newborns to develop normally. As the model for public health-based population genetic screening, newborn screening is recognized nationally as an essential program that aims to ensure the best outcomes for the nation's newborn population.

There are no national newborn screening standards, aside from the *Standard on Blood Collection on Filter Paper* published by the National Committee for Clinical Laboratory Standards[1] and guidance from the Council of Regional Networks for Genetic Services, funded by the Health Resources and Services Administration (HRSA), and limited advice is available from national advisory committees and national medical or public health professional organizations regarding newborn screening policies and conditions to be included in screening mandates. The level of state resources available (personnel, equipment, and service capacity); programs' interpretations of available evidence concerning given conditions (incidence, treatability, and impact); availability or expense of new screening methods; and public advocacy by families, health care professionals, and state legislators have often led to divergence among states regarding which conditions should be mandated for newborn screening. This divergence has resulted in significant disparities in screening services available to infants. In 2000, the American Academy of Pediatrics (AAP) Newborn Screening Task Force[1] indicated that greater uniformity among programs would benefit families, professionals, and public health agencies.

The public health system faces many challenges as newborn screening capabilities continue to evolve. The health care service infrastructure is limited with respect to the interconnections among primary care professionals and subspecialists, particularly in rural areas, a problem complicated by the number and diversity of very rare conditions identified in newborn screening programs. There are geographic limitations in the availability of specific expertise for many of the rare conditions, and considerable needs exist throughout the health care system in the areas of training and education about the disorders detected through newborn screening programs. Furthermore, improvements in the newborn screening system and 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 occur, but they are limited in their ability to ensure long-term management, including the provision of the necessary treatment and services. In addition, new technologies have brought 3 major challenges to newborn screening: (1) expansion of the knowledge base regarding the causes and therefore the treatment or potential treatment of genetic diseases; (2) rapid expansion of diverse technologies, such as multiplex platforms, that may be used in screening; and (3) increased use of tiered testing strategies to enhance the positive predictive value of an initial abnormal result.

The lack of newborn screening program uniformity for infants, the changing dynamics of emerging technology, and the complexity of genetics necessitate assessment of the state of the art in newborn screening and views on future directions such programs could take. In 1999, the AAP Newborn Screening Task Force[2] 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."

In response to this need, the Maternal and Child Health Bureau of HRSA commissioned the American College of Medical Genetics (ACMG) to outline a process of standardization of outcomes and guidelines for state newborn screening programs and to 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 would be definitive and that the subsequent recommendations would be based on the best scientific evidence and analysis of that evidence. ACMG was asked specifically to develop recommendations to address (1) a uniform condition panel (including implementation methods), (2) model policies and procedures for state newborn screening programs (with consideration of a national model), (3) model minimal 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 response to the HRSA/Maternal and Child Health Bureau request.

DEVELOPING A UNIFORM SCREENING PANEL

As indicated, the AAP task force was concerned particularly about the lack of uniformity between the state-based newborn screening programs and the need for "nationally recognized newborn screening system standards and policies." There are few existing systems that allow for the assessment of conditions to determine their

appropriateness for newborn screening. In addition to the original Wilson-Jungner criteria,[3] some states (eg, Nebraska and Washington) have developed such evaluation criteria and systems; other countries (eg, Australia and Belgium) have developed them as well. However, most use criteria that either are difficult to quantify or do not allow conditions to be comparatively ranked adequately. Most are inadequate with respect to the handling of conditions that have similar or overlapping disease markers or may be detected through the use of multiplex technologies but may vary in their analytical and clinical features.

METHODS

Expert Group Development and Process

ACMG convened a group, the Newborn Screening Expert Group, that included participants with expertise in various areas of subspecialty medicine and primary care, health policy, law, ethics, and public health, consumers, and several ad hoc work groups. As an initial step in the process, the expert group developed a set of guiding principles for its work. The establishment of these principals was followed by the development of criteria with which conditions were to be evaluated and the identification of the conditions to be evaluated. A steering committee oversaw the work of this group. The 2 work groups were formed to provide more in-depth analysis in 2 specific areas, that is the uniform panel and its criteria and the diagnosis and follow-up system.

The expert group used a 2-tiered approach for assessing and ranking conditions. In the first tier, with the specific evaluation criteria, conditions were analyzed by recognized experts and other interested individuals to develop a quantification of opinion. In the second tier, the quantification data were subjected to an analysis of the evidence base for each specific screening criterion score. Basic principles developed to guide the decision-making process were factored with the 2 levels of analysis to yield a set of core conditions. Further, additional conditions that are clinically significant that can be revealed during establishment of the diagnosis due to relationships with screening analytes used to identify core conditions or the technology used to screen were identified and referred to as secondary condtions.

Establishing Principles

The following basic principles were developed as a framework for defining the criteria with which to evaluate conditions and to make recommendations. (1) Universal newborn screening is an essential public health responsibility that is critical for improving the health outcomes of affected children. (2) Newborn screening policy development should be driven primarily by the interests of affected newborns, with secondary consideration being given to the interests of unaffected new-

borns, families, health professionals, and the public. (3) Newborn screening is more than testing. It is a coordinated comprehensive system consisting of education, screening, follow-up contact, diagnosis, treatment and management, and program evaluation. (4) The medical home and the public and private components of the screening programs should be in close communication, to ensure confirmation of test results and appropriate follow-up evaluation and care of identified newborns. (5) Recommendations about the appropriateness of conditions for newborn screening should be based on evaluation of scientific evidence and expert opinion. (6) To be included as a primary target condition in a newborn screening program, a condition should meet the following minimal criteria: it can be identified at a time (24–48 hours after birth) at which it would not ordinarily be detected clinically; a test with appropriate sensitivity and specificity is available for it; and there are demonstrated benefits of early detection, timely intervention, and efficacious treatment of the condition. (7) The primary targets of newborn screening should be conditions that meet the criteria listed in principle 6. The newborn screening program also should report any other results of potential clinical significance. (8) Centralized health information data collection is needed for longitudinal assessment of disease-specific screening programs. (9) Total quality management should be applied to newborn 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. (11) Public awareness, coupled with professional training and family education, is a significant program responsibility that must be part of the complete newborn screening system.

Choosing Conditions

The conditions chosen for evaluation were included for ≥1 of several reasons, as follows. They are included in private, state, or national newborn screening programs. They are revealed coincidentally by some of the technologies used in newborn screening. They were identified by members of the expert group as worthy of consideration. They were identified by disease-specific advocacy organizations. They are included in the differential diagnosis of screening results for another condition. In the course of information collection, all conditions were subject to reconsideration. Eighty-four conditions were chosen for consideration.

Developing Evaluation Criteria and Their Comparative Values

The uniform panel working group developed the criteria with which conditions were to be evaluated; these were modified subsequently by the expert group. Criteria were divided into 3 main categories that covered aspects of the condition, that is, (1) clinical characteristics (eg, incidence, burden of disease if not treated, and pheno-

type in the newborn); (2) analytical characteristics of the screening test (eg, availability and features of the platform); and (3) diagnosis, treatment, and management of the condition in acute and chronic forms (this criterion includes the availability of health professionals experienced in diagnosis, treatment, and management).

Within each of these categories, several component criteria were developed (resulting in a total of 19 criteria) for assignment of the comparative value or score. The scoring system recognizes the strengths and limitations found for each condition and summarizes them in a ranking system. Therefore, 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 overruled by scientific evidence of new advances in testing and treatment and should be recognized as opportunities for targeted 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 that have a highly objective scientific basis and others that are associated with more subjective aspects. To the extent possible, the expert group relied on the scientific literature to provide the information on which the recommendations are based. However, some criteria have significant subjective aspects that require the consideration of more than just scientific and expert opinion. For example, issues of cost were considered but were not viewed as central in the analysis of the scientific literature. Cost is an example of a subjective criterion because it is a contextual concern and can be measured only against the value of the outcome.

Collecting Data

The first tier of the analysis was accomplished through the development of a data collection instrument containing the evaluation screening criteria. A survey was conducted to allow for the input of a wide range of individuals and organizations with interest in newborn screening. The data collection instrument included methods not only to collect information from experts but also to quantify that expert opinion regarding features of the conditions under consideration for inclusion in a uniform condition panel.

Before wide distribution, the data collection instrument was pilot tested. Potentially ambiguous language was identified and clarified, and scores were adjusted modestly to reflect the evolving priorities of the expert group. After modification, the data collection instrument was made widely available through passive efforts (eg, Listserv lists of interest groups such as the Genetic Alliance, Association of Public Health Laboratories, and Association of State and Territorial Health Officials) and active efforts (eg, direct approaches to experts on the conditions under evaluation and/or to support groups

for particular conditions under evaluation). In this way, it was possible to acknowledge broad views that were of a more-subjective nature, such as the simplicity of the treatment (parents and individuals with the disorder in question often differed significantly from experts when scoring items such as simplicity of treatment). The results led to a preliminary listing of conditions and their placement in 1 of 3 categories, that is, high scoring, moderately scoring but part of the differential diagnosis of a high-scoring condition, or low scoring and not appropriate for newborn screening at this time. The responses of ≥3 recognized experts for each condition were compared with responses of all respondents regarding that condition, and results were found to be consistent.

Survey results were analyzed statistically. Respondents were characterized to ensure that they were broadly representative of the population. With the recognition that not all who responded have expertise or experience in all aspects of newborn screening for a specific condition, methods were used that allowed data to be aggregated for each criterion for each condition, rather than using the total score for a condition. A mean score for each criterion for each condition was based only on the responses provided for the criterion. Respondents were allowed to insert a "U" if an answer was unknown. The sum of the means was used for the total score assigned to a condition, because the sum of means tends to acknowledge dissenting views more clearly than does the sum of medians.

It is recognized that this relatively open survey process limited the views of experts while considering the views of those less knowledgeable about the individual conditions. However, analyses provided by scientific experts showed that their views were in close agreement with those of most respondents.

Establishing 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. Each condition was considered with respect to the available scientific evidence, such as systematic reviews of reference lists (including Medline, PubMed, and others), books, Internet sources, professional guidelines, clinical evidence, and cost/economic evidence and modeling, for each criterion. The categorization was adjusted in accordance with the evidence. The analysis of the evidence base from the scientific literature included details about the screening tests, the efficacy of treatments, and the adequacy of the knowledge base for the condition. Disease-specific fact sheets were developed to describe this evidence.

At least 2 recognized experts examined the evidence on the fact sheet for all criterion scores for the conditions and assigned a level of evidence for each criterion score,

making the scoring system part of a fuller evidence-based analysis. Therefore, the evaluation of the evidence for the scores in the second tier of analysis is part of a broader assessment of the scientific literature related to the conditions, tests, and treatments. In addition to validating the evidence gleaned from the literature and other sources, the experts assigned a level of quality to the studies from which the evidence was drawn. Adjustments based on the evidence were made primarily on the basis of the accuracy of the information. When significant differences were found between the data collected through the survey and the evidence base, the differences were acknowledged and addressed in each of the fact sheets. Only rarely were adjustments required to align the literature evidence with the views of the survey respondents.

RESULTS

In the first tier of assessment, nearly 300 individuals from the United States and other countries completed the data collection instrument. Many respondents provided information on multiple conditions, yielding information on nearly 4000 individual disease-specific responses. The data are displayed in Table 1 and Fig 1, where the sums of the means are displayed for all conditions. Medium-chain acyl-CoA dehydrogenase deficiency, congenital hypothyroidism, and phenylketonuria were the highest-scoring conditions in this evaluation system, followed by biotinidase deficiency, sickle cell anemia, and congenital adrenal hyperplasia. A number of other conditions that scored in the upper third were also found to have an efficacious treatment and sufficient natural history information to be considered appropriate for newborn screening. Most conditions in the middle third of scores were also included in the differential diagnosis of ≥1 of the higher-scoring conditions. Almost all conditions in the bottom third of scores either lacked a screening test that had been validated in a general newborn population or were deficient in meeting several of the assigned evaluation criteria. Because of limited involvement of infectious disease experts, the expert group chose to defer decision-making on infectious diseases.

A score of 1200 on the data collection instrument was found to provide a logical point of separation between a group of high-scoring conditions (1200–1799 of a possible 2100) and another group of low-scoring (<1000) conditions. A group of conditions with intermediate scores (1000–1199) was identified, all of which were part of the differential diagnosis of a high-scoring core condition but without an efficacious treatment or without a well-understood natural history.

With the use of expert opinion and the validated evidence base, each condition that had been assigned previously to a category on the basis of quantified scores was reconsidered on the basis of the scientific evidence regarding an available screening test, an efficacious treatment, an adequate understanding of the natural history, whether the condition was part of the differential diagnosis of another condition, and whether the screening test results were related to a clinically significant condition. These categories were referred to as the core panel, secondary targets (conditions that are part of the differential diagnosis of a core panel condition), and not appropriate for newborn screening (either no newborn screening test is available or there is poor performance with respect to multiple other evaluation criteria).

DISCUSSION

The basis for decision-making started with whether a screening test is available, which was then overlaid with the overall quantified expert opinion analysis gathered with the data collection information tool. The process of quantifying this expert opinion was informed by literature review and expert validation.

In the first tier of analysis, conditions with scores of >1200 met key criteria and were preliminarily considered appropriate for inclusion in a core newborn screening panel. Conditions scoring <1000 were not considered appropriate for inclusion in the core newborn screening panel at this time. As noted previously, the expert group determined that laboratories should report any result revealed coincidentally in the course of newborn screening that might be clinically significant. In general, the screening test has been optimized for the detection of primary target conditions. Optimizing the technology for a primary target condition does not necessarily optimize the detection of all possible conditions. These conditions are often revealed through diagnostic testing because they are part of the differential diagnosis of a core condition, as occurs with tandem mass spectrometry (MS/MS)-identified cases, but they may be apparent in the screening laboratory because of the technologies used in screening (eg, hemoglobinopathies detected with high-pressure liquid chromatography/isoelectric focusing). Therefore, the expert group designated a category of secondary targets, which included conditions for which results should be made available to health care professionals and/or families by the screening laboratory or for which results are determined during the diagnostic phase of the screening program and provided to families in the course of diagnosis and follow-up care. Most conditions placed in the secondary target category are part of the differential diagnosis of a condition in the core panel. Inclusion in the secondary target category allows for the collection of cases on a national level for additional investigation to understand the disease process and for development of treatment modalities. Regardless of whether programs choose to integrate all such conditions into their broader newborn screening programs, it will be important for them to

TABLE 1   Scores for All Conditions (Sorted in Descending Order of Sum of Mean Scores)

Condition	Abbreviation	Sum of Mean Scores	Percentile
Medium-chain acyl-CoA dehydrogenase deficiency	MCAD	1799	1.00
Congenital hypothyroidism	CH	1718	0.99
Phenylketonuria	PKU	1663	0.98
Neonatal hyperbilirubinemia (kernicterus)	HPRBIL	1584	0.96
Biotinidase deficiency	BIOT	1566	0.95
Sickle cell anemia (hemoglobin SS disease)	Hb SS	1542	0.94
Congenital adrenal hyperplasia (21-hydroxylase deficiency)	CAH	1533	0.93
Isovaleric acidemia	IVA	1493	0.89
Very long-chain acyl-CoA dehydrogenase deficiency	VLCAD	1493	0.89
Maple syrup disease	MSUD	1493	0.89
Classic galactosemia	GALT	1473	0.88
Hemoglobin S/β-thalassemia	Hb S/βTh	1455	0.87
Hemoglobin S/C disease	Hb S/C	1453	0.86
Long-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency	LCHAD	1445	0.84
Glutaric acidemia type I	GA I	1435	0.83
3-Hydroxy-3-methyglutaric aciduria	HMG	1420	0.82
Trifunctional protein deficiency	TFP	1418	0.81
Multiple carboxylase deficiency	MCD	1386	0.80
Benign hyperphenylalaninemia	H-PHE	1365	0.78
Methylmalonic acidemia (mutase deficiency)	MUT	1358	0.77
Homocystinuria (attributable to cystathionine β-synthase deficiency)	HCY	1357	0.76
3-Methylcrotonyl-CoA carboxylase deficiency	3MCC	1355	0.75
Hearing loss	HEAR	1354	0.73
Methylmalonic acidemia (Cbl A,B)	Cbl A,B	1343	0.72
Propionic acidemia	PROP	1333	0.71
Carnitine uptake defect	CUD	1309	0.69
Galactokinase deficiency	GALK	1286	0.69
Glucose-6-phosphate dehydrogenase deficiency	G6PD	1286	0.67
β-Ketothiolase deficiency	BKT	1282	0.66
Citrullinemia	CIT	1266	0.65
Argininosuccinic acidemia	ASA	1263	0.64
Tyrosinemia type I	TYR I	1257	0.63
Short-chain acyl-CoA dehydrogenase deficiency	SCAD	1252	0.61
Tyrosinemia type II	TYR II	1249	0.60
Glutaric acidemia type II	GA2	1224	0.59
Medium/short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency	M/SCHAD	1223	0.58
Cystic fibrosis	CF	1200	0.57
Variant hemoglobinopathies (including hemoglobin E)	Var Hb	1199	0.55
Human HIV infection	HIV	1193	0.54
Defects of biopterin cofactor biosynthesis	BIOPT(BS)	1174	0.53
Medium-chain ketoacyl-CoA thiolase deficiency	MCKAT	1170	0.52
Carnitine palmitoyltransferase II deficiency	CPT II	1169	0.51
Methylmalonic acidemia (Cbl C,D)	Cbl C,D	1166	0.49
Argininemia	ARG	1151	0.48
Tyrosinemia type III	TYR III	1149	0.47
Defects of biopterin cofactor regeneration	BIOPT(REG)	1146	0.46
Malonic acidemia	MAL	1143	0.45
Carnitine/acylcarnitine translocase deficiency	CACT	1141	0.43
Isobutyryl-CoA dehydrogenase deficiency	IBG	1134	0.42
2-Methyl-3-hydroxybutyric aciduria	2M3HBA	1132	0.41
Carnitine palmitoyltransferase I deficiency (liver)	CPT IA	1131	0.40
2-Methylbutyryl-CoA dehydrogenase deficiency	2MBG	1124	0.39
Hypermethioninemia	MET	1121	0.37
Dienoyl-CoA reductase deficiency	DE RED	1119	0.36
Galactose epimerase deficiency	GALE	1066	0.35
3-Methylglutaconic aciduria	3MGA	1057	0.34
Severe combined immunodeficiency	SCID	1047	0.33
Congenital toxoplasmosis	TOXO	1041	0.31
Familial hypercholesterolemia (heterozygote)	FHC	1038	0.30
Carnitine palmitoyltransferase I deficiency (muscle)	CPT IB	1009	0.29
Citrullinemia type II	CIT II	1001	0.28

TABLE 1 Continued

Condition	Abbreviation	Sum of Mean Scores	Percentile
Ornithine transcarbamylase deficiency	OTC	942	0.27
Guanidinoacetate methyltransferase deficiency	GAMT	922	0.24
Wilson disease	WD	922	0.24
Diabetes mellitus, insulin dependent	IDDM	891	0.23
Neuroblastoma	NB	864	0.22
Arginine:glycine amidinotransferase deficiency	AGAT	861	0.20
Turner syndrome	TURNER	847	0.19
Adenosine deaminase deficiency	ADA	841	0.18
Carbamoylphosphate synthetase deficiency	CPS	833	0.17
α1-Antitrypsin deficiency	A1AT	819	0.16
Congenital cytomegalovirus infection	CMV	779	0.14
Duchenne and Becker muscular dystrophy	DMD	776	0.12
Fragile X syndrome	FX	776	0.12
Congenital disorder of glycosylation type Ib	CDG Ib	766	0.11
Smith-Lemli-Opitz syndrome	SLO	759	0.10
Biliary atresia	BIL	744	0.08
Hurler-Scheie syndrome	MPS-1H	707	0.07
X-linked adrenoleukodystrophy	ALD	705	0.06
Fabry disease	FABRY	661	0.05
Creatine transport defect	CR TRANS	646	0.04
Lysosomal storage diseases	LSD	638	0.02
Pompe disease	POMPE	613	0.01
Krabbe disease	KRABBE	447	0.00

have the diagnostic confirmatory results for all such cases, because the results have a direct impact on the calculation of false-positive rates of screening for the core panel conditions.

After conditions were preliminarily categorized on the basis of their data collection instrument scores, the evidence base, as reflected in fact sheets developed for each condition, was assessed. If a clinically significant condition in the core panel did not have the scientific evidence to support the availability of an efficacious treatment, then it was moved to the secondary target category. Similarly, if it was determined that an understanding of the natural history of the condition was insufficient to justify primary screening, then the condition was moved to the secondary target category. When test results identified carriers of the conditions definitively, the handling of carrier information was moved into the secondary target category.

Figure 2 demonstrates the decision-making algorithm. It is important to note that the algorithm presumes an ongoing review of conditions to determine their continued or newly identified appropriateness for newborn screening as new tests and treatments evolve. The data collection instrument used in this project provides an assessment of only one aspect of a broader decision-making process required for establishing a newborn screening uniform panel. An ongoing analysis of the scientific evidence must be overlaid on the quantified expert opinion.

Clearly, the first decision to screen is based on the availability of a sensitive specific screening test that can be performed in the 24- to 48-hour period after birth. A total of 29 conditions are considered appropriate for newborn screening because they have a screening test, an efficacious treatment, and adequate knowledge of natural history (Table 2). The conditions best meeting all of the criteria established by the expert group are medium-chain acyl-CoA dehydrogenase, congenital hypothyroidism, and phenylketonuria. Among conditions assigned to the core panel are 9 organic acidurias, 6 amino acidurias, 5 disorders of fatty acid oxidation, 3 hemoglobinopathies associated with a hemoglobin S allele, and 6 other conditions. Twenty-three of the 29 conditions in the core panel are identified with multiplex technologies such as MS/MS.

On the basis of the evidence, 6 of the 35 conditions placed initially in the core panel were moved into the secondary target category, which expanded to 25 conditions that are part of the differential diagnosis of a core panel condition. Knowledge of these secondary targets (ie, from newborn screening or follow-up test results) can be clinically important to the family. In addition to the 54 conditions identified in Table 2, the expert group identified 27 conditions that were not considered appropriate for newborn screening, either because they met few evaluation criteria or because they lacked a screening test.

There were limitations. Conditions with limited evidence reported in the scientific literature were more difficult to evaluate with the data collection instrument. For example, some conditions have been reported for
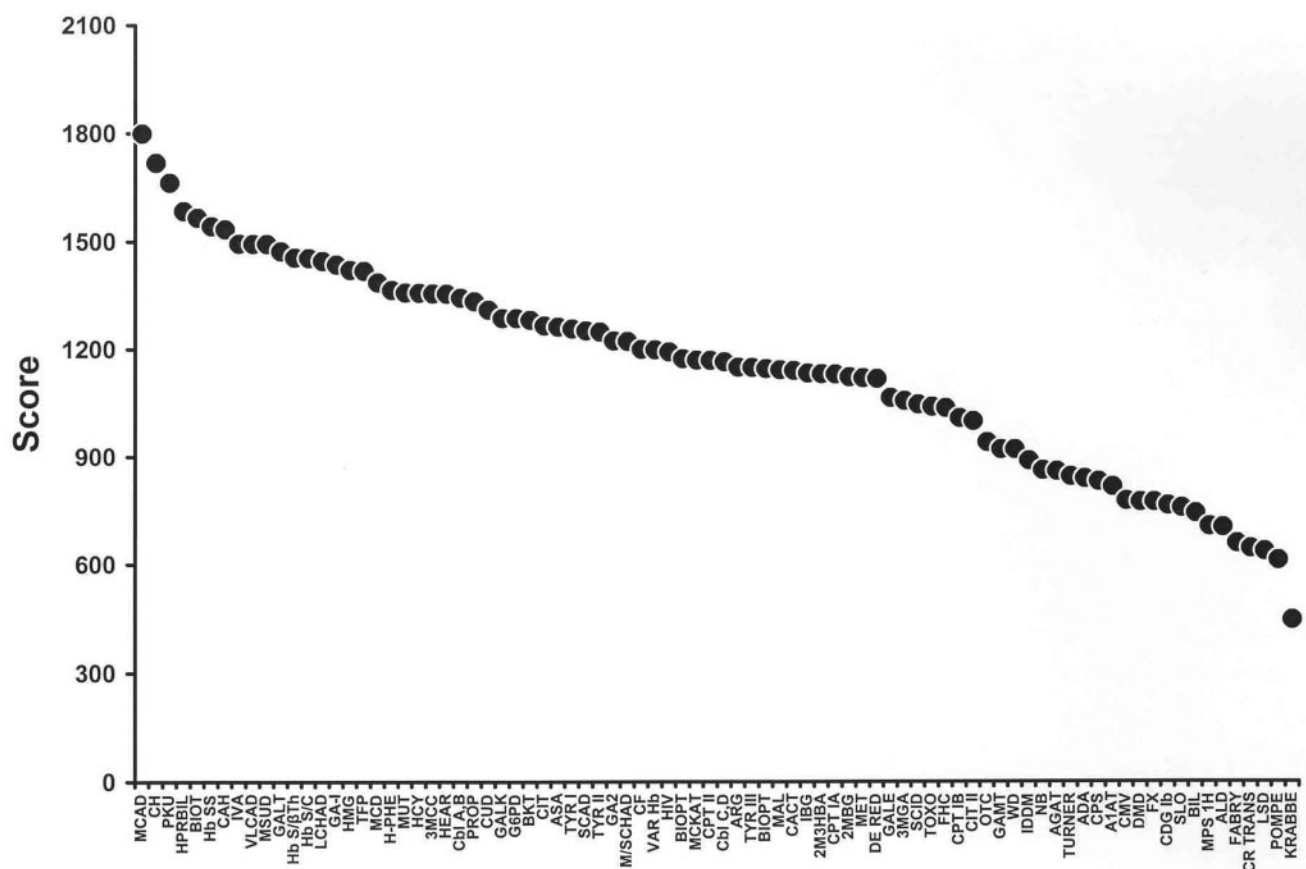
FIGURE 1
Scoring according to test availability. This scoring separates conditions that have an acceptable, validated, population-based screening test from those that do not. Abbreviations for conditions are as listed in Table 1.

≤10 families in the world. Many conditions were found to occur in multiple forms, distinguished by age of onset, severity, or other features. Furthermore, unless a condition was already included in newborn screening programs, a potential for bias was apparent in the information related to some criteria. The power of the statistical analyses and the blending of 2 forms of evaluation also presented limitations. The data collection process in the first tier of the analysis was limited also by the significant variability in the numbers of individuals responding for the different conditions. Because of limitations in the scientific evidence for these rare diseases, there was significant reliance on the opinions of experts on the conditions. There were many conditions that scored close to other conditions, and it is unlikely that the statistical power provided in these analyses was sufficient to discriminate accurately among the conditions in a ranking system. Nevertheless, groups of scores were assessed, and natural separations between groups became apparent. In such circumstances, expert opinion, with reasoning that applied first principles of genetic medicine to the evidence and to the quality of the data, determined the placement of the conditions in particular categories.

PROGRAM EVALUATION, COST-EFFECTIVENESS, AND FUTURE NEEDS OF THE NEWBORN SCREENING SYSTEM

The Newborn Screening System

Because the appropriate functioning of the system is critical to realizing improved outcomes, the components of a screening system were examined by the expert group during the project. (Information was obtained from program reports submitted to the National Newborn Screening and Genetics Resource Center and is based on information available as of October 2003.) The goal of the evaluation was to determine the extent to which states have addressed the many aspects of components of this system and to recommend performance standards to improve the quality of the system. The ability to ensure appropriate diagnosis and management is considered to be primarily a system responsibility. Limitations and significant variability were identified in the components of prenatal education, screening, follow-up services, diagnosis, treatment, and program management. For example, financing across state and county lines is constrained by state-based Medicaid rules; service delivery is fragmented on a categorical or disease basis; there is insufficient support to bridge geo-
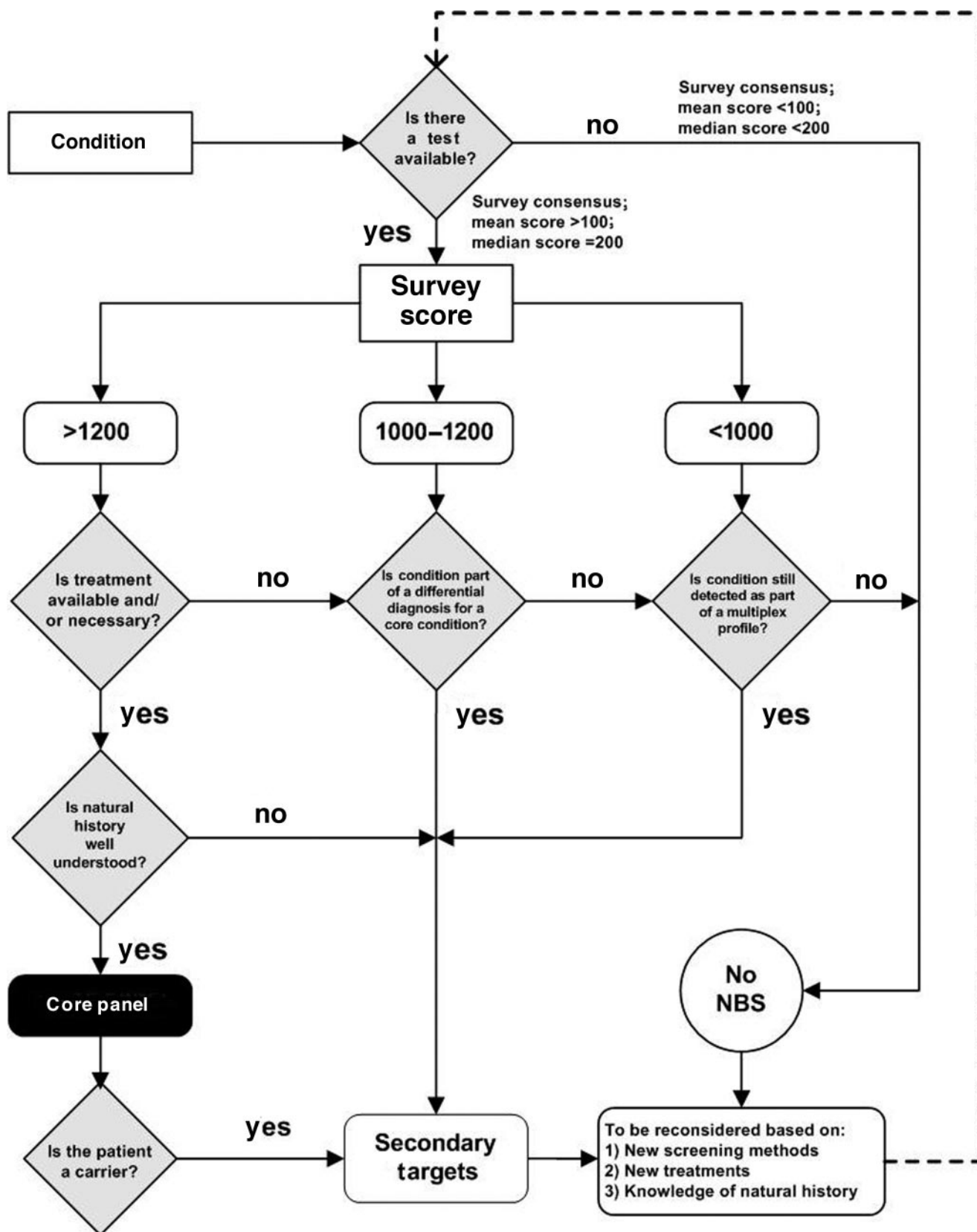
FIGURE 2
Condition evaluation and decision-making algorithm. NBS indicates newborn screening.

TABLE 2 Core Panel and Secondary Targets in Newborn Screening Panel

	OA	FAO	AA	Hemoglobinopathies	Other
Core panel					
	IVA	MCAD	PKU	Hb SS[a]	CH
	GA I	VLCAD	MSUD	Hb S/βTh[a]	BIOT
	HMG	LCHAD	HCY[a]	Hb S/C[a]	CAH[a]
	MCD	TFP	CIT		GALT
	MUT[a]	CUD	ASA		HEAR
	3MCC[a]		TYR I[a]		CF
	Cbl A,B[a]				
	PROP				
	BKT				
Secondary targets					
	Cbl C,D[a]	SCAD	H-PHE	Var Hb[a]	GALK[a]
	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		

OA indicates disorders of organic acid metabolism; FAO, disorders of fatty acid metabolism; AA, disorders of amino acid metabolism. Abbreviations used for conditions are listed in Table 1.
[a] Conditions for which specific discussions of unique issues are found in the text.

graphic barriers; it is difficult to identify experienced health care professionals for complex care (eg, centers of excellence for genital reconstructive surgery for congenital adrenal hyperplasia or confirmation of metabolic diagnoses); there is misinterpretation of privacy regulations (eg, the Health Information Portability and Accountability Act); there is underuse and lack of uniformity of information technology; collaborative management and care are often constrained by systems of reimbursement for services; state sovereignty sometimes dictates individual approaches; and there is variability in financing of screening programs.[4]

There are national and state roles in addressing these limitations, and states must retain their significant roles and responsibilities. They have 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 comprehensive financing of services. Service coordination at state and local levels must be considered, as well as program integration with the State Children's Health Insurance Program, early intervention programs, Title V programs, and similar services.

It is apparent, however, that all state programs could benefit from a more-robust national role in newborn screening. Because so many of the conditions screened for among 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 de-

fining core condition panels. This would allow states to apply the best science to their own considerations when determining their roles in expanded screening.

Practice guidelines also could be developed at a national level by interested organizations. The expert group identified a clear gap between the information available and the information needed by primary care professionals to facilitate an immediate response in the event of a screen-positive case. In response, the expert group developed an action sheet for each core condition and secondary target, to facilitate immediate responses on the part of primary care professionals with respect to the expected steps in diagnosis and follow-up care.

There are also potentially expanded national roles in oversight, data collection, and program evaluation, as well as development of educational materials to support newborn screening. Depending on the overall incidence of particular conditions, regional collaborative groups such as those funded by HRSA could 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 based largely 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 management, 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 to ensure that health care professionals are available locally, regionally, and nationally will be important to ensuring access to high-quality services.

Cost-Effectiveness Analysis

A basic cost-effectiveness assessment project was performed to inform the decision-making process. The assessment 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, because costs often are the basis on which such decisions are made.

Costs and benefits related to screening for particular conditions or groups of conditions were evaluated after mapping them over major disease outcomes (eg, life expectancy, cerebral palsy/stroke, seizures, developmental delay, hearing loss, and vision loss). Costs were obtained from the literature, and benefits were determined from expected outcomes with and without early treatment or intervention. The results of these analyses indicated that most newborn screening programs improve outcomes and reduce overall costs. Furthermore, technologies such as MS/MS or high-pressure liquid

chromatography save money because of their multiplexing capabilities and low screening false-positive rates. The identification of potentially affected individuals at such an early age leads to many years over which the benefits accrue and aggregate over costs.

CONCLUSIONS

Significant variability in the conditions for which newborns are screened led to this project to assess the scientific and medical evidence and the views of various individuals and interest groups associated with the conditions being considered. Throughout this undertaking, scientific literature and expert opinion formed the basis for information collection and assessment. The expert panel considered a range of information, from disease-specific information to the full breadth of the newborn screening system, in evaluating 84 conditions. There was an effort to overlay the evidence, where available, on expert opinion. The process of quantifying this expert opinion was informed by literature review and expert validation. It is important to acknowledge that there was limited scientific evidence available on the rare disorders considered by the expert panel. Furthermore, because there was limited activity in the area of coordinated data collection and analysis, it seemed unlikely that robust scientific evidence would be available in the near future. Therefore, reliance on experts and their ability to apply first principles[5,6] was required.

Guiding principals for newborn screening and criteria were established for evaluating conditions. The conditions being considered were assigned initially, through expert analysis, to 1 of 3 categories, depending on how they met the screening criteria. The categories were core panel, secondary targets (conditions that are part of the differential diagnosis of a core panel condition), and not appropriate for newborn screening (either no newborn screening test is available or there is poor performance with respect to multiple other evaluation criteria). Each condition was then evaluated to determine the extent to which the scientific evidence supports the availability of a test and a treatment, whether the natural history of the condition is well understood, and whether the information provided by testing indicates the possible presence of the condition or of a carrier state.

The expert panel identified 29 conditions for which screening should be mandated. An additional 25 conditions were identified because they are part of the differential diagnosis of a condition in the core panel, they are clinically significant and revealed with screening technology but lack an efficacious treatment (eg, some identified with MS/MS technology), or they represent incidental findings for which there is potential clinical significance (hemoglobinopathies). The expert group thought it was important that such findings be communicated to the health care service community and to families. In addition, the view that the technologies used in newborn screening should be maximized is inherent in the recommendation that all clinically significant information discovered through newborn screening should be provided to the relevant health care professionals and/or the family. The expert group recommends that state newborn screening programs mandate screening for all core panel conditions defined in this article; mandate reporting of all secondary target conditions defined herein and reporting of any abnormal results that may be associated with clinically significant conditions, including definitive identification of carrier status; maximize the use of multiplex technologies; and consider that the range of benefits realized through newborn screening includes treatments that go beyond an infant's death or morbidity.

The full breadth of the newborn screening system was assessed, including a brief review of its cost-effectiveness. Numerous barriers to implementation of an optimal screening and follow-up program were identified. Recommended actions to overcome these barriers include establishment of a national role in the scientific evaluation of conditions and the technologies with 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.

The recommendations are as follows. (1) Programs should continue to improve the components of the system beyond the initial screening, communicate results, and ensure that affected newborns enter short-term follow-up care. (2) Reporting procedures should be standardized. (3) Reports of confirmatory results should be obtained. (4) There should be improved oversight (eg, Joint Commission on Accreditation of Hospital Organizations) of hospital-based screening activities, to improve tracking of screen-positive cases. (5) There should be more uniformity in the definition of the performance standards (eg, repeat test versus second test) monitored and reported by programs. (6) The quality assurance programs involving the diagnostic and follow-up system should be enhanced. (7) National oversight and authority, with appropriate resources, should be provided. (8) Systems should be in place for collection of data about individuals identified as screen-positive in newborn screening programs.

ACKNOWLEDGMENTS

The Newborn Screening Expert Group was as follows: Newborn Screening Steering Committee: Jose Cordero, MD, MPH, Centers for Disease Control and Prevention; E. Steven Edwards, MD, Past President, AAP; R. Rodney

Howell, MD, ACMG, University of Miami School of Medicine; Jennifer L. Howse, PhD, President, March of Dimes Birth Defects Organization; Michele A. Lloyd-Puryear, MD, PhD, HRSA; Marie Y. Mann, MD, MPH (project officer), HRSA; Tricia Mullaley, BA, Genetic Alliance; Peter van Dyck, MD, MPH, Maternal and Child Health Bureau; HRSA; Michael S. Watson, PhD (project director), ACMG; Newborn Screening Expert Group: R. Rodney Howell, MD (chair), ACMG, University of Miami School of Medicine; William Becker, DO, Association of Public Health Laboratories, Ohio State Department of Health, Ohio State University; Coleen Boyle, PhD (ex officio), Centers for Disease Control and Prevention; George C. Cunningham, MD, MPH, Genetics Disease Branch, California Department of Health Services; Michael R. DeBaun, MD, MPH, Washington University School of Medicine; Stephen M. Downs, MD, Indiana University School of Medicine; Edward Goldman, JD, University of Michigan Hospitals and Health System; Stephen I. Goodman, MD, University of Colorado Health Sciences Center; Fernando Guerra, MD, MPH, San Antonio Metropolitan Health District; W. Harry Hannon, PhD (ex officio), Centers for Disease Control and Prevention; James Hanson, MD (ex officio), National Institute of Child Health and Human Development, National Institutes of Health; Cecilia Larson, MD, New England Newborn Screening Program; Michele A. Lloyd-Puryear, MD, PhD (ex officio), HRSA; Marie Y. Mann, MD, MPH (ex officio) (project officer), HRSA; Scott McLean, MD, LTC, MC (ex officio), United States Army; Gurvaneet Randhawa, MD, MPH (ex officio), Agency for Healthcare Research and Quality; Piero Rinaldo, MD, PhD, Mayo Clinic College of Medicine; Derek Robertson, MBA, JD, health care consultant/family representative; Mark Rothstein, JD, University of Louisville; Robert D. Steiner, MD, Oregon Health and Science University; Sonia Suter, MS, JD, George Washington University Law School; Bradford Therrell, PhD, National Newborn Screening and Genetics Resource Center, University of Texas Health Science Center at San Antonio; Thomas Tonniges, MD, AAP; Wanda Yazzie, RN, MPH, New Mexico Department of Health; Michael S. Watson, PhD (project director), ACMG; Newborn Screening Conditions and Criteria Work Group: Piero Rinaldo, MD, PhD (chair), Mayo Clinic College of Medicine; Donald Bailey, PhD (family representative), University of North Carolina at Chapel Hill; Celia I. Kaye, PhD, MD, University of Texas Health Science Center at San Antonio; Alex R. Kemper, MD, MPH, University of Michigan Health System; Michele A. Lloyd-Puryear, MD, PhD, HRSA; Marie Y. Mann, MD, MPH (project officer), HRSA; Kenneth Pass, PhD, Association of Public Health Laboratories, New York State Department of Health; Jennifer M. Puck, MD, National Human Genome Research Institute, National Institutes of Health; Bradford Therrell, PhD, National Newborn

Screening and Genetics Resource Center, University of Texas Health Science Center at San Antonio; Michael S. Watson, PhD (project director), ACMG; Newborn Screening External Review Group: Franklin Desposito, MD, University of Medicine and Dentistry of New Jersey; Gary Hoffman, BS, Wisconsin State Laboratory of Hygiene; Kathy Stagni, BS, Organic Acidemia Association; Arnold Strauss, MD, Vanderbilt University School of Medicine; Tracy Trotter, MD, private pediatric practice; Anne M. Willey, PhD, JD, Association of Public Health Laboratories, New York State Department of Health; Newborn Screening Diagnosis and Follow-up Work Group: Harvey L. Levy, MD (chair), Children's Hospital Boston, Harvard Medical School; James R. Eckman, MD, Emory University School of Medicine; Michele A. Lloyd-Puryear, MD, PhD, HRSA; Fred Lorey, PhD, California Department of Health Services; Marie Y. Mann, MD, MPH (project officer), HRSA; Deborah L. Marsden, MBBS, Children's Hospital Boston, Harvard Medical School; Julie Miller, BS, Nebraska Department of Health; Danielle Laraque, MD, Mount Sinai School of Medicine; Kelly R. Leight, JD (family representative), CARES Foundation; Derek Robertson, MBA, JD, health care consultant/family representative; Bradford Therrell, PhD, National Newborn Screening and Genetics Resource Center, University of Texas Health Science Center at San Antonio; Michael S. Watson, PhD (project director), ACMG; Barbara Yawn, MD, American Academy of Family Physicians; Newborn Screening Health Information Portability and Accountability Act Work Group Edward B. Goldman, JD (chair), University of Michigan Hospitals and Health System; Coleen Boyle, PhD, Centers for Disease Control and Prevention; Beverly Dozier, PhD, Department of Health and Human Services; Lynn D. Fleisher, PhD, JD, Sidley, Austin, Brown, and Wood; Mark Rothstein, JD, University of Louisville; Sonia Suter, MS, JD, George Washington University Law School.

REFERENCES

1. Clinical Laboratory Standards Institute. *Blood Collection on Filter Paper for Newborn Screening Programs: Approved Standard.* 4th ed. Wayne, PA: CLSI; 2003
2. American Academy of Pediatrics, Newborn Screening Task Force. Serving the family from birth to the medical home: newborn screening: a blueprint for the future. *Pediatrics.* 2000;106:389–427
3. National Academy of Sciences. *Genetic Screening: Programs, Principles, and Research.* Washington, DC: National Academy of Sciences; 1975
4. US Congress, Office of Technology Assessment. Data and methods used in OTA's cost-effectiveness analysis of strategies for newborn screening. In: *Healthy Children: Investing in the Future.* Washington, DC: US Government Printing Office; 1988:236–241. Publication OTA-H-345. Available at: www.wws.princeton.edu/ota/ns20/alpha_f.html
5. McCabe ERB. Principle of newborn screening for metabolic disease. *Perinatol Neonatol.* 1982;6:63–73
6. Therrell BL, Panny SR, Davidson A, et al. US newborn screening system guidelines: statement of the Council of Regional Networks for Genetic Services. *Screening.* 1992;1:135–147