

Decision analysis, economic evaluation, and newborn screening: challenges and opportunities

Lisa A. Prosser, MS, PhD¹, Scott D. Grosse, PhD², Alex R. Kemper, MD, MPH³, Beth A. Tarini, MD, MS¹ and James M. Perrin, MD⁴

The number of conditions included in newborn screening panels has increased rapidly in the United States during the past decade, and many more conditions are under consideration for addition to state panels. The rare nature of candidate conditions for newborn screening makes their evaluation challenging. The scarcity of data on the costs of screening, follow-up, treatment, and long-term disability must be addressed to improve the evaluation process for nominated conditions. Decision analyses and economic evaluations can help inform policy decisions for newborn screening programs by providing a systematic approach to synthesizing available evidence and providing

projected estimates of long-term clinical and economic outcomes when long-term data are not available. In this review, we outline the types of data required for the development of decision analysis and cost-effectiveness models for newborn screening programs and discuss the challenges faced when applying these methods in the arena of newborn screening to help inform policy decisions.

Genet Med 2012;14(8):703–712

Key Words: cost-effectiveness; decision analysis; economic evaluation; health policy; newborn screening

The U.S. Department of Health and Human Services Secretary's Advisory Committee on Heritable Disorders in Newborns and Children (SACHDNC) makes recommendations for conditions to be included in the recommended uniform newborn screening panel.¹ The activities of the Advisory Committee are supported by an external evidence review workgroup. This external evidence review group conducts systematic reviews summarizing available information on the benefits and harms of screening for nominated conditions.^{1–6} As part of this process, the evidence review group has conducted several reviews of candidate disorders. Although the group searched for cost-effectiveness analyses or sources of evidence that could be used to develop decision analysis or cost-effectiveness models of screening for these disorders, rarely were such studies identified.^{1–6} Certain coauthors (L.A.P., A.R.K., J.M.P.) participate in this evidence review group. Other coauthors have previously addressed the advantages and challenges to using cost-effectiveness analysis to inform newborn screening policy decisions (S.D.G.)^{7,8} and to address ethical issues relating to newborn screening (B.A.T.).^{9,10}

This review describes the types of data and data elements required to develop decision analytic models and conduct economic evaluations of newborn screening programs. This review is intended as a resource for researchers designing studies to evaluate the clinical and economic outcomes for newborn screening programs and to aid reviewers of decision analytic

models and economic evaluations of newborn screening programs. Decision analysis can provide an approach for synthesizing evidence from disparate sources to assist decision makers in estimating potential long-term health benefits and harms. Cost-effectiveness analysis can provide information on the relative value of screening for a condition or set of conditions.

DECISION ANALYTIC MODELING

Decision analysis is a systematic approach to decision making under conditions of uncertainty that has been applied to clinical and public health problems.¹¹ Decision analytic models can be used to simulate randomized clinical trials for new health interventions, to project beyond the clinical trial time frame, or to compare treatment protocols not directly compared in head-to-head trials. The decision analytic approach allows the decision maker to identify which alternative is expected to yield the most health benefit. It can also allow analysts to characterize the uncertainty associated with projections of clinical and economic outcomes over the long term,¹² which is important given the lack of long-term outcomes data for most conditions considered for newborn screening.

To develop a decision analytic model, each aspect of the decision problem requires attention: defining of the set of possible alternative methods for disorder identification (e.g., universal screening and clinical identification), screening technologies or protocols, possible choices regarding timing of implementation

¹Child Health Evaluation and Research Unit, Division of General Pediatrics, University of Michigan, Ann Arbor, Michigan, USA; ²National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ³Department of Pediatrics, Duke University, Durham, North Carolina, USA; ⁴Center for Child and Adolescent Health Policy, MassGeneral Hospital for Children, Harvard Medical School, Boston, Massachusetts, USA. Correspondence: Lisa A. Prosser (lisapros@med.umich.edu)

Submitted 13 June 2011; accepted 27 January 2012; advance online publication 5 April 2012. doi:10.1038/gim.2012.24

(i.e., screening at birth or at a later age), uncertainties (e.g., consideration of long-term health outcomes), health outcomes (e.g., disability, death), probabilities of these identified outcomes, and the values assigned to each health outcome.¹³ The decision analytic modeling approach uses evidence from all available sources, such as clinical trials, cohort studies, observational studies, case-control studies, meta-analyses, and expert opinion; synthesizes the evidence; and can account for the strength of each evidence source by including the range of uncertainty associated with each parameter input.¹⁴

For newborn screening policy decisions, a decision analytic approach can leverage existing data for clinical and economic outcomes with unpublished data and estimates based on expert opinion to provide policy-relevant information. Given the rare nature of screened conditions, the evidence base to evaluate screening for such conditions is typically sparse. A decision analytic modeling approach can be valuable to decision makers by providing estimates of long-term outcomes and in identifying the parameters, when varied over ranges identified to reflect the uncertainty associated with a parameter input, that have the greatest impact on results. By helping identify projected outcomes and key data gaps, decision analysis can supplement the available evidence base and also help prioritize further research areas.¹³ This paper provides a brief introduction to the terms of decision analysis; more detailed primers are available elsewhere.¹⁵⁻¹⁹

ECONOMIC EVALUATION

If a decision analytic model incorporates costs, it becomes an economic evaluation model (Table 1). The two main types of economic evaluations in health care are cost-effectiveness analysis and cost-benefit analysis. This article focuses on cost-effectiveness analysis, as it is more commonly used for evaluating health interventions. Cost-benefit analysis, which requires the conversion of health benefits into monetary terms, has been less well-accepted by the medical community, with the exception of environmental health applications for which the use of cost-benefit analysis is more common.²⁰

Cost-effectiveness analysis is used to measure the relative value of health-care interventions in terms of the cost per health benefit gained, such as the cost per child identified with a newborn-screened condition or the cost per death averted.²¹ If health outcomes are measured using a preference-based

measure, such as quality-adjusted life years, which integrate morbidity and mortality, then the analysis is considered to be a cost-utility analysis, a special case of cost-effectiveness analysis.²⁰⁻²² Quality-adjusted life years are calculated by multiplying the value for each health state, called a “health utility,” by the duration of the health state.²³ The “health utility” is scaled between 1.0, which represents perfect health, and 0.0, which represents a health state equivalent to being dead, although health states can be assigned a value less than zero, which represents a state of health considered as being “worse than dead.”²²

For newborn screening applications, cost-effectiveness is calculated by dividing the net costs by the net health benefits of newborn screening for a disorder, or set of screened disorders, as compared with clinical identification of the specified disorder(s), assuming that both the numerator and denominator are positive. If the health denominator is negative, the potential harms of the screening program outweigh projected health benefits and screening would be considered to be “dominated” by clinical identification, and no ratio is calculated. If the numerator of net costs is negative, the intervention is said to be cost saving, and no ratio is calculated. If net costs are positive, screening does not reduce total costs, but can still be considered cost-effective. Most health interventions are not cost saving, but many are cost-effective.^{24,25}

Screening for most heritable disorders is not cost saving and requires a net investment in resources. Exceptions are disorders, such as phenylketonuria, that have a low mortality rate and a high lifetime cost of treatment for complications associated with late diagnosis. For a given test cost, a lower prevalence of a condition is associated with higher cost per case detected and lower cost-effectiveness. For that reason, multiplex testing, e.g., tandem mass spectrometry, is an important strategy to improve the cost-effectiveness of screening for low-prevalence conditions. Whether screening is considered cost-effective can vary depending on the threshold used to define cost-effectiveness.^{26,27}

TYPES OF DATA NEEDED FOR DECISION ANALYTIC MODELS AND ECONOMIC EVALUATIONS

Required data elements fall into three general categories: clinical outcomes (measures and associated probabilities), values, and costs. Figure 1 shows a simplified schematic of a decision

Table 1 Types of decision analysis and economic evaluations

Type	Measurement of resource use	Measurement of health benefits	Description of health benefits
Decision analysis	None	Health outcomes, including QALYs	Clinical end points or QALYs
Cost-effectiveness analysis	Dollars	Health outcomes	Clinical end points, such as cases averted, hospitalizations averted, or deaths averted
Cost-utility analysis	Dollars	QALYs	QALYs incorporate morbidity and mortality effects into a single metric
Cost-benefit analysis	Dollars	Dollars	Dollars (typically measured via willingness-to-pay survey questions)

QALYs, quality-adjusted life years.

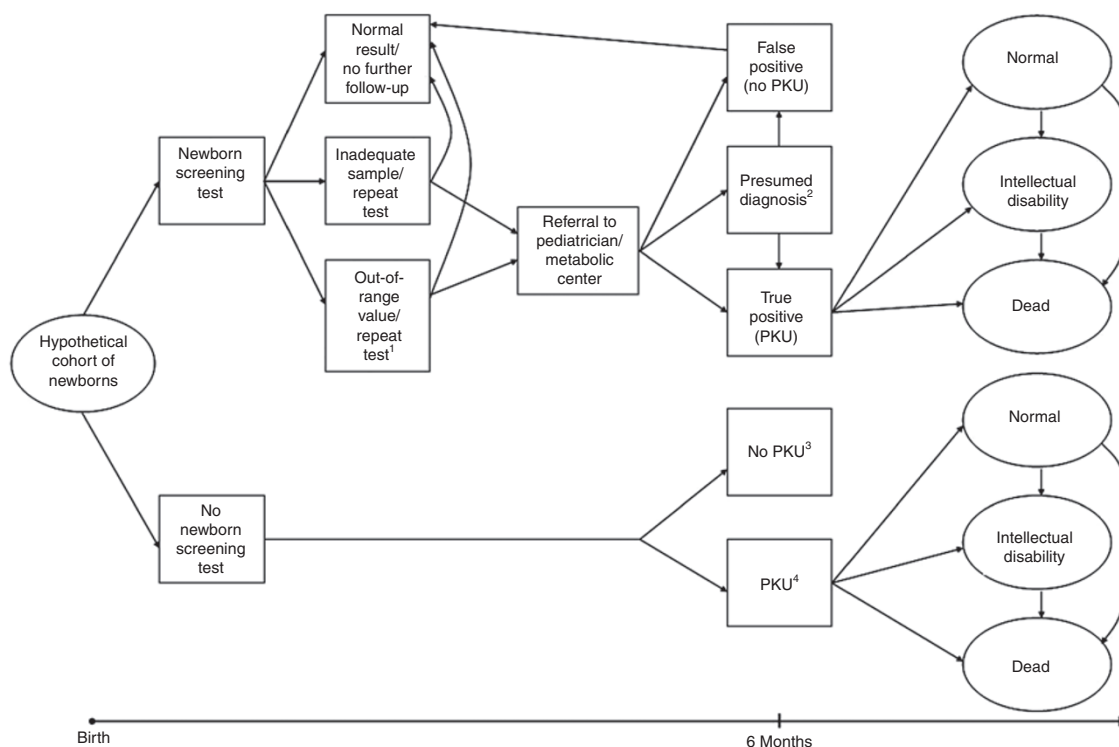


Figure 1 Simplified schematic of simulation model to project clinical and economic outcomes for PKU using one of many possible screening algorithms. ¹Or “positive screen”; ²based on additional follow-up, newborns with initial presumed diagnoses are eventually categorized as either “false positive” or “true positive”; ³children without PKU follow natural history of general population; ⁴diagnosed through clinical identification. PKU, phenylketonuria.

analytic model. Each health state included in the decision analysis is assigned both a probability and a value. This value, also known as a health utility weight, is used to reflect the quality of life associated with a health state (described in more detail in the following). For a cost-effectiveness analysis, each health state or transition to a health state would also be associated with a cost.

Screening and clinical outcomes

Screening phase. Characteristics of screening algorithms include sensitivity and specificity of screening protocols and the accuracy of confirmatory testing relative to diagnostic evaluation. Reliable data on the sensitivity and specificity of alternative screening protocols, such as universal or targeted screening, are needed as inputs for accurate decision analytic models and cost-effectiveness analyses. It is also important to characterize detection rates for conditions in the absence of screening.

Although newborn screening experts generally seek to maximize sensitivity to avoid missed cases, the false-positive rate can significantly affect costs and therefore is relevant in evaluating cost-effectiveness. As specificity increases, the number of false-positive test results decreases. Therefore, a highly specific screening protocol can improve cost-effectiveness by reducing the number of children who require short-term follow-up and confirmatory testing but are not ultimately diagnosed with the disorder in question. For example, the introduction of tandem mass spectrometry for phenylketonuria helped reduce costs by

increasing test specificity and consequently reduced the number of false-positives and associated costs.²⁸ Specificity typically has substantially more influence on costs than does sensitivity because of the low incidence of screened conditions. Lower specificity is associated with higher costs of follow-up for false-positives, which can be sizeable for a screening test with a large number of false-positive results for each true-positive case identified. Suppose that sensitivity is reduced from 99.9% to 95%; this would reduce identified cases, and associated health outcomes, by ~5%. In contrast, reducing specificity from 99.9% to 95% would raise the number of false-positives to 50 times the original estimate and the costs of follow-up would rise proportionally.

Medical evaluation phase. Each condition requires specification of a protocol for follow-up evaluation that defines sequences of confirmatory or diagnostic testing for different out-of-range test results. The results of the diagnostic evaluation may be either dichotomous or tiered, e.g., presumptive positive, possible, or negative. For purposes of economic evaluation, the services required for each stage of confirmatory testing need to be specified as well as the probabilities of true-positives at each stage.

Clinical outcomes for individuals identified via screening versus clinical identification. Defining outcomes for the screening program, clinical identification in the absence of

screening, and the probability of these outcomes requires data on long-term health outcomes, such as data on hospitalization, cognitive function, disability, and mortality. One potential bias when measuring screening and clinical outcomes is a failure to adjust for differences in the spectrum of severity of cases detected clinically as compared with those detected by screening. For many disorders, such as congenital hypothyroidism, medium-chain acyl-CoA dehydrogenase (MCAD) deficiency, and cystic fibrosis, screening identifies many children with relatively mild phenotypes who would have been less likely to be identified in the absence of screening in addition to those with severe phenotypes who would have been detected clinically in the absence of screening.^{29–31} For these disorders, using data on the severity of outcomes among clinically identified cases to project outcomes in screened cohorts in the absence of screening will overstate the magnitude of morbidity and disability prevented by screening.³² An appropriate model design for this situation would allow for differing levels of severity in screened and clinically identified cohorts. On the other hand, cases of sudden death due to a disorder, such as congenital adrenal hyperplasia (CAH) or MCAD deficiency, are likely to be missed in the absence of screening.^{33,34} Whether the net effect of both types of bias leads to an underestimate or overestimate of the magnitude of mortality risk is dependent on the particular disorder in question.

One particularly challenging task for the development of a decision analytic model for newborn screening is the need for accurate data on unscreened cohorts. Using historical data on unscreened cohorts who may not have had access to currently available treatments could be misleading and result in substantial overestimates of the benefits of screening. An appropriate evaluation of a screening program requires that comparable treatments were available to both screened and clinically identified cohorts. Such data are rarely available, with notable exceptions for cystic fibrosis and MCAD deficiency.³⁵

For example, a long-term study of outcomes in birth cohorts with MCAD deficiency in Australia during 1973 to 2002 demonstrated that ascertainment in unscreened cohorts improved over time as a result of increased clinical awareness, and the severity of outcomes diminished with increased clinical awareness.³⁶ In particular, for births during 1995 to 2002, there were no cases of intellectual disability in unscreened cohorts with MCAD deficiency.³⁶ Also, the authors adjusted downward the estimates of the frequency of mortality among clinically diagnosed children in unscreened cohorts born during 1994 to 2002 to reflect a lower risk of death in asymptomatic cases missed without screening.³⁷ At the population level, the net effect of ascertainment bias outweighed the opposite bias of missed diagnoses of sudden death among undiagnosed children.³⁵

Another example of the implications of using historical data on unscreened cohorts is CAH. The key outcome used to evaluate the effectiveness of newborn screening for CAH is prevention of death associated with adrenal crises. Previous analyses of screening for CAH relied on clinical data from populations without adequate treatment.³⁸ Available data from

high-income countries indicate a low level of mortality associated with unscreened CAH, even taking into account the missed cases.^{33,39} As a consequence, economic evaluations that use historical estimates of mortality associated with unscreened CAH will likely overestimate the cost-effectiveness of screening for CAH in such countries.³⁸ On the other hand, screening has benefits besides prevention of death, such as the prevention of morbidity.

Effectiveness of treatment for identified individuals. Estimates of treatment effectiveness for both short- and long-term clinical end points are needed. Long-term outcomes are of key importance for defining the effectiveness of an intervention and will, as a result, be of key importance for determining cost-effectiveness. For example in 1999, Denmark introduced newborn screening for congenital toxoplasmosis based on evidence of favorable short-term outcomes associated with treatment. However, in 2007, after data on long-term outcomes revealed no evidence of lasting benefit, Denmark discontinued screening for congenital toxoplasmosis.⁴⁰ Estimates of long-term adherence and adverse events associated with recommended treatments are required and should include the full spectrum of possible outcomes, along with their associated costs and intended and unintended consequences.

Cost inputs

In the United States, newborn screening programs are public health programs funded at the state level. For most public programs, the societal perspective is the most appropriate analytic perspective to assume for an economic evaluation because this perspective is the most comprehensive analytic perspective and will include all direct medical costs, direct nonmedical costs (e.g., transportation), and opportunity costs (e.g., patient or family time costs associated with screening, follow-up, and care).²¹ This review assumes the use of the societal perspective for costs. Costs can be separated into several subcategories: those relating to the costs of screening, treatment, and short- and long-term costs of care for the identified condition (**Table 2**). The costs of a newborn screening program will include the downstream costs of care as well as the costs of the initial screen, follow-up testing, and diagnosis.

Screening phase. The initial screening test represents only a subset of the total costs associated with a newborn screening program. The costs associated with this screening phase involve more than the cost of performing the initial screening test. It also includes the costs associated with reporting positive or uncertain results and the collection of repeat specimens and repeated screens, if necessary.

Currently, there are few published data on the costs of newborn screening programs. Each state has its own set of screened conditions and processes for conducting screening. States vary in the number of specimens collected per infant, with 12 states routinely collecting and testing two specimens for each infant. States vary greatly in the extent to which they fund follow-up

Table 2 Categories and definitions for required data inputs

Category	Subcategory	Specific components	Description/examples
Clinical outcomes (probabilities)	Screening phase	Screening outcomes	Data on sensitivity, specificity, and duration of time for following up test results
	Medical evaluation phase and clinical outcomes	Natural history	Epidemiologic data for short- and long-term outcomes for the condition identified via clinical identification or with newborn screening; includes incidence of condition, mortality rates
Treatment characteristics		Estimates of the effectiveness of treatment for short- and long-term health outcomes; adherence rates; adverse events	
Costs	Screening phase	Test costs	Cost of the initial screen
	Medical evaluation phase	Costs of following up out-of-range test result	All costs associated with medical evaluation following an out-of-range test result
	Treatment	Direct medical costs	Hospitalizations, outpatient visits, drug treatments, procedures, diagnostic tests, medical equipment, other costs
Direct nonmedical costs		Transportation costs, special education, home modifications, other costs	
Opportunity costs		Patient time for testing and treatment, informal caregiver time	
Valuation of health outcomes	Screening results	Screening results	Public or parent values for false-positive results as valued using QALYs
		Health outcomes	Public or parent/patient values for short- and long-term health outcomes included in the natural history model as valued using QALYs
	Treatments	Treatments	Loss in health-related quality of life associated with treatment regimens measured using QALYs. Loss in health-related quality of life could include difficulty of adhering to dietary treatments or losses in relation to painful and difficult transplant procedures.
		Treatment-related adverse events	Loss in health-related quality of life for adverse events associated with recommended treatment regimens measured using QALYs

QALYs, quality-adjusted life years.

testing, particularly long-term follow-up, as well as genetic counseling and cascade testing of family members. The substantial variation across states and sharing of resources across other public health programs can make tracking the costs specific to a newborn screening program difficult. Also, testing costs vary according to the annual number of specimens tested in a laboratory because of economies of scale, which can result in more than a threefold difference in average testing costs for a given disorder.³⁸ Commonly, cost-effectiveness analyses assign testing cost based on data from just one state, which does not adequately reflect variability across states, therefore limiting the ability to draw generalizable conclusions.

Medical evaluation phase. When a positive newborn screening result occurs, a number of costs accrue related both to additional testing and to medical evaluation of the infant. For example, an infant with a positive result will undergo confirmatory or diagnostic testing and medical evaluations by one or more physicians (e.g., primary-care, specialist) depending upon the complexity of the workup or the emergent nature of the evaluation. These costs will include direct medical costs for diagnostic tests and clinician fees as well as time and transportation costs for patients and their families. Many economic analyses oversimplify this process by assuming fixed costs for positive specimens. In reality, the type of tests ordered and the urgency with which the infant is brought in for testing, both of which affect costs, often vary based on the amount by which

the test result exceeds the screening cutoff and will vary by disorder.

Costs of care for individuals identified via screening or clinical identification. Downstream costs of screening include the net costs of medical care and treatment for a screened individual as compared with what the costs would have been in the absence of screening. Direct medical costs for hospitalizations, procedures, drug treatments, outpatient visits, medical equipment, and rehabilitation may differ. For example, additional costs of caring for patients with MCAD deficiency detected through screening could include the costs of additional medical visits, preventive hospitalizations to avoid fasting for MCAD deficiency infants with other illnesses, possible costs of carnitine supplementation, and parent time costs. Preventive treatment could reduce costs associated with treating metabolic crises and their disabling sequelae, including special education and caregiving costs. Whether net medical costs associated with screening for a given disorder are positive or negative is difficult to reliably predict. In the case of MCAD deficiency, an Australian study found little overall difference in costs of hospitalizations whether or not screening was performed,⁴¹ contrary to other published economic analyses.⁴²

For patients identified with severe combined immunodeficiency, costs of caring for identified children would include costs associated with the receipt of interventions such as hematopoietic stem cell transplant as well as costs of possible

adverse events associated with hematopoietic stem cell transplant and costs of infections if treatment is not perfectly effective.⁴³ For severe combined immunodeficiency patients identified through clinical identification, categories of costs would largely be the same, including costs of hematopoietic stem cell transplant, adverse events, and infections, but the number of infections and their associated morbidity and mortality would be greater under clinical identification because of the lower effectiveness of hematopoietic stem cell transplant in preventing infections among clinically identified individuals.

Valuation of health outcomes

With newborn screening, changes in health-related quality of life can be associated with screening test results, differences in health states for the screened disorder, and recommended treatments, such as transplants or dietary restrictions. For an economic evaluation, health utilities should be assigned to each health outcome. Both direct and indirect methods are available for valuing health outcomes.⁴⁴

The valuation of health outcomes for newborn-screened conditions presents a number of challenges for researchers. Methodological challenges of valuing children's health states include the need for proxy respondents, lack of validated methods for valuing health in young children, and the need for the inclusion of spillover effects on family members' quality of life.^{45,46} Another challenge for valuing health outcomes related to newborn-screened conditions is the lack of data on long-term outcomes, such as the effect of a condition on employment or educational attainment. For many conditions, information on long-term health outcomes is scarce, which makes it difficult to assign values to outcomes that are not well described. The rarity of the conditions makes it difficult to employ indirect methods of valuation, in which a patient (or in this case the parent proxy) would rate the condition using a predefined set of health attributes.

This specific set of challenges has resulted in substantial variability in health utilities used in existing cost-utility analyses of newborn screening programs. A recent review of cost-utility analyses of newborn screening for metabolic disorders found a high level of variability in the specific weights assigned to the same conditions.⁴⁷ For example, serious intellectual disability was assigned weights in the different studies ranging from 0.06 (equivalent to being close to death) to 0.67.⁴⁷

Conventionally, only the loss of health utility for the affected individual is included in quality-adjusted life year estimates, but the inclusion of spillover effects on the quality of life of other family members is gaining recognition as an important component of economic evaluations, especially for childhood health conditions.⁴⁵ Economic evaluations of children's health should consider the relevance of family spillover effects, defined as the loss in health-related quality of life for a parent or caregiver due to a child's condition. We are aware of only one newborn screening economic evaluation that has incorporated family spillover effects by asking parents to calculate losses in health-related quality of life for both their children and themselves.⁴⁸

In practical terms, the dominant sources of quality-adjusted life year gains from newborn screening are from the prevention of morbidity and mortality in the newborn. Future research should focus on improving consistency and accuracy of measuring health outcomes for newborn-screened conditions, regardless of whether family spillover effects are included.

POLICY-RELEVANT OUTCOMES

A decision analytic model of newborn screening strategies can provide short- and long-term estimates of the outcomes important to policy decisions for newborn screening programs. Policy makers have considered outcomes, such as expected numbers of infant and child deaths prevented, cases of permanent disability avoided, and changes in health-care costs, in the criteria that have been used to assess newborn screening tests in the United States.⁴⁹ Decision analytic models can also provide projections of testing-related outcomes, such as expected numbers of positive and false-positive screens, and utilization-related outcomes, including expected hospitalizations, procedures, and outpatient visits, as well as the cost-effectiveness of alternative screening strategies. This comprehensive set of testing-related outcomes is likely to be valuable in weighing the evidence for policy decisions and newborn screening.

CHALLENGES AND OPPORTUNITIES

Conducting high-quality decision analyses and economic evaluations requires high-quality evidence for all of the aforementioned areas: health outcomes, costs, and quality of life. Data limitations can be challenging for any health application, given the many categories of evidence required, but these data limitations are particularly salient for newborn screening candidate disorders due to the low incidence and the long time frame over which outcomes need to be considered. Constructing a decision analytic model and assembling the necessary inputs can be resource intensive and time consuming, which can present an additional challenge when the need for a policy decision is urgent.

Lack of data on long-term outcomes

The key data challenge for measuring health outcomes is the absence of data on long-term outcomes of newborn screening programs. More data are becoming available for short-term outcomes of newborn screening programs, such as the sensitivity and specificity of screening protocols. As long-term follow-up programs become established, the increasing availability of long-term data for screened conditions can potentially help inform decisions about candidate conditions that share similar characteristics. However, such research efforts are not likely to address the unique challenge of assessing what long-term outcomes would be in the absence of screening. The advantage of a decision analytic modeling approach is that a range of assumptions for outcomes of clinically identified cohorts can be explored in the analysis.

Decision analysis represents a promising approach to evaluating newborn screening policy options. The use of decision

analytic models can assist decision makers by providing estimates of health benefits and possible risks for varying time horizons and for varying assumptions for test characteristics, treatment benefits, and possible harms. Understanding the ranges of possible outcomes for different input assumptions can be informative to decision makers, given the absence of long-term data for most conditions that are nominated for newborn screening. A decision analysis can consider a range of assumptions for key issues such as a broader spectrum of disease detected by screening or potential harms of treatment.

Difficulties in defining costs

There are numerous challenges to obtaining a full account of the costs associated with a newborn screening program. Most of the available data are from the health-care system perspective. However, because state-level newborn screening programs are public programs, the appropriate perspective to use in most cases would be the societal perspective. Some analyses have used the payer perspective; however, given the substantial burden of many of these conditions on family members, alternative perspectives such as the payer perspective could result in substantial underestimation of the burden of illness and associated benefits of averted illness. In these more narrow analytic perspectives, some of the types of costs listed earlier, such as patient or family time costs, would be excluded.²¹ An additional challenge to the collection of accurate cost data in the United States is the fragmented structure of health-care financing in which costs are covered by various payers, including state public health programs, public and private health payers, and the family.

Even more challenging is the estimation of costs for clinically identified cases. Because identification and treatment may have improved over time, the use of historical data for clinically identified cases may be misleading. For the comparator strategy of clinical identification, the appropriate approach for an economic evaluation should assume usual care from the same time period. Treatment patterns are likely to represent a substantial improvement when compared to historical data from prior to the initiation of newborn screening. For example, the two classic examples of newborn screening programs that are cost saving are phenylketonuria and congenital hypothyroidism.⁵⁰ However, the magnitude of reduction in costs has likely been overstated in published economic evaluations because of widespread misinterpretation of available data on long-term outcomes in unscreened cohorts for these two disorders. Long-term outcomes for patients with late-treated phenylketonuria show that the degree of cognitive impairment on average is less than was assumed in previously published economic evaluations.⁵⁰

Identifying data for the comparator strategy

The comparator strategy is the alternative against which a new screening policy is compared. Choice of the comparator can affect conclusions about the cost-effectiveness of screening. For example, the comparator strategy for universal screening could

be either targeted screening or no screening. Targeted screening is generally difficult to use as a comparator because of a lack of information about the effectiveness and costs of targeted versus universal screening. However, considering alternative screening strategies is one advantage of using a decision analytic approach, allowing for the consideration of alternatives for which little data are available.

Valuation of health outcomes

The valuation of health outcomes using health utilities for newborn screening presents methodological challenges due to a lack of standardization regarding optimal approaches for assigning health utilities to child health outcomes.^{45,47} Improved methods for valuing children's health conditions needs to be paired with better characterization of long-term outcomes to provide inputs appropriate for decision analytic modeling and economic evaluation.

Defining the scope of the analysis

Calculating the cost-effectiveness of screening for a single condition may not be straightforward. If the out-of-range value could be associated with more than one condition, then it may be more appropriate to evaluate the cost-effectiveness of screening for the panel of conditions instead of the single condition being added to the panel and then evaluating the incremental cost-effectiveness of the expanded panel as compared with the original panel. These other conditions are referred to as "secondary targets." This situation is not unique to newborn screening and is analogous to other screening protocols for which the reported results can include incidental findings unrelated to the original screening condition. If these findings are reported and followed up on, these must also be included in the cost-effectiveness analysis. In the case of newborn screening, this requires the analyst to explicitly and carefully define the scope of the analysis.

FUTURE CONSIDERATIONS

Decision analyses and cost-effectiveness analyses of newborn screening programs require data on short-term screening results and long-term outcomes. Analyses of other types of screening programs, such as mammography for breast cancer or colonoscopy for colon cancer, can provide instruction for the analysis of newborn screening programs. Such previous analyses of screening and public health programs demonstrate the value of decision analytic modeling to help inform clinical and public health decisions for which long-term data are not available.¹² In the case of newborn screening, parameters that are characterized using limited data can be assessed for their relative impact on outcomes using sensitivity analysis; a decision analytic approach could help prioritize research areas by identifying which parameters have the most effect on projected outcomes.¹³ Over the long term, the use of decision analytic modeling, along with increased primary data collection, could optimize the use of existing data on newborn screening programs. Decision analytic modeling also provides

an opportunity to model alternative strategies in addition to universal screening and clinical identification, such as targeted screening, and can provide results on alternative strategies based on a combination of the best available evidence in the absence of direct clinical trials.

Specific findings from decision analytic models of newborn screening can provide useful insights. For example, a recent analysis of newborn screening for MCAD deficiency demonstrated that assumptions about the loss in quality of life associated with dietary treatment were potentially influential on cost-effectiveness results,⁴⁸ and this finding could potentially apply to other conditions.

The collection of long-term data on health outcomes, costs, and quality of life should be incorporated into current efforts to create registries and collect data for conditions identifiable through newborn screening, including those on the recommended universal newborn screening panel. States and regional collaboratives can play a role in collecting these data. For example, the National Newborn Screening Translational Research Network is developing standards for how such data can be efficiently collected through a standard database platform. The Network is testing this approach in pilot studies of screening for lysosomal storage disorders. In addition to outcomes data from a screened cohort, it is necessary to have comparable information on long-term outcomes of a representative unscreened cohort for the development of a decision analytic model. This information may be derived either from population-based surveillance in populations without screening or from retrospective screening studies of stored dried blood specimens collected prior to the initiation of screening for the disorder(s) of interest.⁵¹ Both study designs have been applied in the case of MCAD deficiency.^{37,52} The creation and maintenance of registries and related data collection efforts will require resources; ongoing funding from state and federal agencies will be needed to support these efforts.

The role of decision analysis in the context of newborn screening for rare conditions is likely to differ from the application of decision analysis to more common conditions. The scarcity of data, particularly for long-term outcomes, will likely result in a much greater reliance on expert opinion for the development of modeling inputs and in greater uncertainty for modeling results. Despite the greater uncertainty that is likely to be associated with newborn screening simulation models, results from these models can still play an important role in providing a range of possible benefits and harms associated with screening alternatives. For example, a model could provide an estimate of the range of cases prevented, deaths prevented, and/or number of children requiring treatment, as well as other health outcomes, for universal screening compared to clinical ascertainment. Estimating plausible ranges for even a small set of key outcomes could provide useful context for clinical and policy decisions.

The role that cost-effectiveness evidence will play in newborn screening policy is still evolving, given that there has been so little evidence to date. The charge of the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children

currently includes cost-effectiveness analysis as one category of evidence to be considered by the Committee, but it is only one of several criteria considered by the Committee. The threshold for determining whether or not an intervention is cost-effective is not clearly identified in the United States and will likely vary with characteristics of the intervention and target population.^{26,27} More broadly, the role of cost-effectiveness analysis in the context of the Affordable Care Act is unclear.⁵³ In similar policy contexts, such as the consideration of new vaccines by the Advisory Committee on Immunization Practices, cost-effectiveness evidence is also one of the categories of evidence considered in policy decisions.

SUMMARY

This review summarizes the approach and data needs for conducting decision analytic modeling and economic evaluations of newborn screening programs. In addition to data needs that have been identified for long-term follow-up,⁵⁴ this review identifies additional areas to be considered for primary data collection in long-term studies of newborn-screened conditions, such as in the design of prospective cohort studies, and other large-scale data collection efforts. As newborn screening programs continue to expand, the collection of long-term data on newborn-screened conditions could be valuable for informing the evaluation of new candidate conditions that share characteristics of currently screened conditions. In addition to collection of long-term data on health outcomes, the scarcity of data on the costs of screening, follow-up, treatment, and long-term disability must be addressed to improve the evaluation process for nominated conditions. Decision analyses and economic evaluations can help inform policy decisions for newborn screening programs. However, current data limitations have restricted the availability of high-quality evaluations for currently screened and candidate conditions.

ACKNOWLEDGMENTS

This work was supported by subcontracts to the University of Michigan, Duke University, and MassGeneral Hospital for Children under prime contract (HHSP23320045014XI) to the Altarum Institute from the Department of Health and Human Services, the Health Resources and Services Administration, and the Maternal and Child Health Bureau. We also appreciate the expert research assistance of Kara Lamarand.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Perrin JM, Knapp AA, Browning MF, et al. An evidence development process for newborn screening. *Genet Med* 2010;12:131–134.
2. Kemper AR, Browning M. Evidence Review: Pompe Disease. 2008. <http://www.hrsa.gov/heritabledisorderscommittee/reports/evidenceReviewPompeOct2008.htm>.
3. Knapp AA, Kemper AR, Perrin JM. Evidence Review: Krabbe Disease. 2009. <http://www.hrsa.gov/heritabledisorderscommittee/reports/PerrinLettertoCommittee10-23-2009.htm>.

4. Knapp AA, Metterville DR, Kemper AR, Perrin JM. Evidence Review: Hemoglobin H Disease. 2010. <http://www.hrsa.gov/heritabledisorderscommittee/reports/hemogolbinh.pdf>.
5. Knapp AA, Metterville DR, Kemper AR, Prosser LA, Perrin JM. Evidence Review: Critical Congenital Cyanotic Heart Disease. 2010. <http://www.hrsa.gov/heritabledisorderscommittee/reports/CCCHDEvidenceReview.pdf>.
6. Lipstein EA, Vorono S, Browning MF, et al. Systematic evidence review of newborn screening and treatment of severe combined immunodeficiency. *Pediatrics* 2010;125:e1226–e1235.
7. Grosse SD. Cost-effectiveness as a criterion for newborn screening policy decisions: a critical review. In: Baily MA, Murray T (eds). *Ethics and Newborn Genetic Screening: New Technologies, New Forces, New Challenges*. Johns Hopkins University Press: Baltimore, 2008:58–88.
8. Grosse SD. Economic evaluations of newborn screening interventions. In: Ungar WJ (ed). *Economic Evaluation in Child Health*. Oxford University Press: New York, NY, 2010:113–132.
9. Tarini BA, Goldenberg A, Singer D, Clark SJ, Butchart A, Davis MM. Not without my permission: parents' willingness to permit use of newborn screening samples for research. *Public Health Genomics* 2010;13:125–130.
10. Tarini BA, Burke W, Scott CR, Wilfond BS. Waiving informed consent in newborn screening research: balancing social value and respect. *Am J Med Genet C Semin Med Genet* 2008;148C:23–30.
11. Haddix AC, Teutsch SM, Corso PS. *Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation*, 2nd edn. Oxford University Press: Oxford, New York, 2003.
12. Goldie SJ, Corso PS. Decision analysis. In: Haddix AC, Teutsch SM, Corso P (eds). *Prevention Effectiveness*, 2nd edn. Oxford University Press: New York, 2003:103–126.
13. Weinstein MC. *Clinical Decision Analysis*. Saunders: Philadelphia, 1980.
14. Hunink MGM, Glasziou PP, Siegel JE, et al. *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge University Press: New York, 2001.
15. Naimark D, Krahn MD, Naglie G, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 5—Working with Markov processes. *Med Decis Making* 1997;17:152–159.
16. Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 4—Analyzing the model and interpreting the results. *Med Decis Making* 1997;17:142–151.
17. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 3—Estimating probabilities and utilities. *Med Decis Making* 1997;17:136–141.
18. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on medical decision analysis: Part 2—Building a tree. *Med Decis Making* 1997;17:126–135.
19. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis: Part 1—Getting started. *Med Decis Making* 1997;17:123–125.
20. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edn. Oxford University Press: New York, 2005.
21. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. Oxford University Press: New York, NY, 1996.
22. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Pub Health* 2000;21:1–25.
23. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1–30.
24. Cohen JT, Neumann PJ, Weinstein MC. Does preventive care save money? Health economics and the presidential candidates. *N Engl J Med* 2008;358:661–663.
25. Stone PW, Teutsch S, Chapman RH, Bell C, Goldie SJ, Neumann PJ. Cost-utility analyses of clinical preventive services: Published ratios, 1976–1997. *American Journal of Preventive Medicine* 2000;19(1):15–23.
26. Bridges JF, Onukwugha E, Mullins CD. Healthcare rationing by proxy: cost-effectiveness analysis and the misuse of the \$50,000 threshold in the US. *Pharmacoeconomics* 2010;28:175–184.
27. Grosse SD. Assessing Cost Effectiveness in Health Care: The History of the \$50,000 per QALY Threshold. *Pharmacoeconomics and Outcomes Research* 2008;8(2):165–178.
28. Pandor A, Eastham J, Chilcott J, Paisley S, Beverley C. Economics of tandem mass spectrometry screening of neonatal inherited disorders. *Int J Technol Assess Health Care* 2006;22:321–326.
29. Alm J, Hagenfeldt L, Larsson A, Lundberg K. Incidence of congenital hypothyroidism: retrospective study of neonatal laboratory screening versus clinical symptoms as indicators leading to diagnosis. *Br Med J (Clin Res Ed)* 1984;289:1171–1175.
30. Carpenter K, Wiley V, Sim KG, Heath D, Wilcken B. Evaluation of newborn screening for medium chain acyl-CoA dehydrogenase deficiency in 275 000 babies. *Arch Dis Child Fetal Neonatal Ed* 2001;85:F105–F109.
31. Comeau AM, Parad RB, Dorkin HL, et al. Population-based newborn screening for genetic disorders when multiple mutation DNA testing is incorporated: a cystic fibrosis newborn screening model demonstrating increased sensitivity but more carrier detections. *Pediatrics* 2004;113:1573–1581.
32. Grosse SD, Van Vliet G. Prevention of intellectual disability through screening for congenital hypothyroidism: how much and at what level? *Arch Dis Child* 2011;96:374–379.
33. Strnadová KA, Votava F, Lebl J, et al. Prevalence of congenital adrenal hyperplasia among sudden infant death in the Czech Republic and Austria. *Eur J Pediatr* 2007;166:1–4.
34. Dott M, Chace D, Fierro M, et al. Metabolic disorders detectable by tandem mass spectrometry and unexpected early childhood mortality: a population-based study. *Am J Med Genet A* 2006;140:837–842.
35. Grosse SD. Assessing the clinical utility of newborn screening. In: Khoury MJ, Bedrosian S, Gwinn M, Higgins J, Ioannidis JP, Little J (eds). *Human Genome Epidemiology*, 2nd edn. Oxford University Press: New York, 2009:517–532.
36. Wilcken B, Hammond J, Siilink M. Morbidity and mortality in medium chain acyl coenzyme A dehydrogenase deficiency. *Arch Dis Child* 1994;70:410–412.
37. Wilcken B, Haas M, Joy P, et al. Outcome of neonatal screening for medium-chain acyl-CoA dehydrogenase deficiency in Australia: a cohort study. *Lancet* 2007;369:37–42.
38. Yoo BK, Grosse SD. The cost effectiveness of screening newborns for congenital adrenal hyperplasia. *Public Health Genomics* 2009;12:67–72.
39. Grosse SD, Van Vliet G. How many deaths can be prevented by newborn screening for congenital adrenal hyperplasia? *Horm Res* 2007;67:284–291.
40. Roser D, Nielsen HV, Petersen E, Saugmann-Jensen P, Norgaard-Pedersen PB. Congenital toxoplasmosis—a report on the Danish neonatal screening programme 1999–2007. *J Inherit Metab Dis* 2010;33(suppl 2):S241–S247.
41. Haas M, Chaplin M, Joy P, Wiley V, Black C, Wilcken B. Healthcare use and costs of medium-chain acyl-CoA dehydrogenase deficiency in Australia: screening versus no screening. *J Pediatr* 2007;151:121–126, 126.e1.
42. Carroll AE, Downs SM. Comprehensive cost-utility analysis of newborn screening strategies. *Pediatrics* 2006;117(5 Pt 2):S287–S295.
43. McGhee SA, Stiehm ER, McCabe ER. Potential costs and benefits of newborn screening for severe combined immunodeficiency. *J Pediatr* 2005;147:603–608.
44. Brazier J, ebrary Inc. *Measuring and valuing health benefits for economic evaluation*. Oxford University Press: Oxford; New York, 2007.
45. Prosser LA, Hammit JK, Keren R. Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations. *Pharmacoeconomics* 2007;25:713–726.
46. Petrou S. Should health gains by children be given the same value as health gains by adults in an economic evaluation framework? In: Ungar WJ (ed). *Economic Evaluation in Child Health*: Oxford, New York, NY, 2010:271–287.
47. Grosse SD, Prosser LA, Asakawa K, Feeny D. Use of QALY Weights for Neurosensory Impairments in Cost-Utility Analyses of Early Childhood Vaccines and Newborn Metabolic Screening: A Critique. *Expert Review of Pharmacoeconomics & Outcomes Research* 2010;10(3):293–308.
48. Prosser LA, Kong CY, Rusinak D, Waisbren SL. Projected costs, risks, and benefits of expanded newborn screening for MCADD. *Pediatrics* 2010;125:e286–e294.
49. Watson MS, Mann MY, Lloyd-Puryear MA, Rinaldo P, Howell RR, American College of Medical Genetics Newborn Screening Expert Group. Newborn screening: Toward a uniform screening panel and system - Executive summary. *Pediatrics* 2006;117(5):S296–S307.
50. Grosse SD. Late-treated phenylketonuria and partial reversibility of intellectual impairment. *Child Dev* 2010;81:200–211.
51. Grosse SD. Assessing the clinical utility of newborn screening. In: Khoury MJ (ed). *Human Genome Epidemiology: Building the Evidence for Using Genetic Information to Improve Health and Prevent Disease*, 2nd edn. Oxford University Press: Oxford; New York, 2010:517–532.

52. Pourfarzam M, Morris A, Appleton M, Craft A, Bartlett K. Neonatal screening for medium-chain acyl-CoA dehydrogenase deficiency. *Lancet* 2001;358:1063–1064.
53. Sox HC. Comparative effectiveness research: a progress report. *Ann Intern Med* 2010;153:469–472.
54. Kemper AR, Boyle CA, Aceves J, et al. Long-term follow-up after diagnosis resulting from newborn screening: statement of the US Secretary of Health and Human Services' Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children. *Genet Med* 2008;10:259–261.