Report and Recommendations of the Pilot Studies Workgroup

To the Advisory Committee on Heritable Diseases in Newborns and Children

May, 2016

Jeffrey R. Botkin, M.D., M.P.H.
Don Bailey, Ph.D., M.Ed.
Anne Marie Comeau, Ph.D.
Carla Cuthbert, Ph.D.
Kellie B. Kelm, Ph.D.
Dietrich Matern, M.D., Ph.D.
Mark McCann
Scott M. Shone, Ph.D.
Mike Watson, MS, Ph.D.
Tiina Urv, Ph.D.

I. Introduction

Newborn bloodspot screening is conducted through state-based public health programs to enable early detection and treatment of a broad range of conditions. Until approximately 2006, there was substantial variation across state programs in both the conditions targeted and laboratory methods used. Furthermore, there had been a long-standing concern in the field that some conditions were being added to state panels without an adequate evidence-base.(1) These challenges were addressed, in part, through the creation in 2004 of the Secretary's Advisory Committee on Heritable Diseases in Newborns and Children (SACHDNC), now referred to as the Advisory Committee on Heritable Disorders in Newborns and Children (ACHDNC).(2) The SACHDNC Charter states that, among other responsibilities, it shall:

- make systematic evidence-based and peer-reviewed recommendations that include the heritable disorders that have the potential to significantly impact public health for which all newborns should be screened…

The ACHDNC has developed an evidence review process for making recommendations to the Secretary of DHHS regarding the addition of conditions (or the removal of conditions from) the Recommended Uniform Screening Panel (RUSP).[3] State programs use the determinations of the Secretary to guide decisions about conditions targeted by their programs. This approach has been highly effective in reducing state-to-state variability in both the conditions targeted and the testing approaches used.

An evidence-based decision-making process is dependent on quality data. In this context, evidence must be available on the analytic validity of the proposed screening test and follow-up diagnostic tests, the clinical validity of the screening/diagnostic testing, and the clinical utility of population-based screening in order to ensure appropriate treatment or preventive interventions. Both the potential benefits and harms of the screening effort must be considered.[4] Data on these elements of a screening program come from pilot studies.

"Pilot studies" have been defined in various ways throughout the development of newborn screening (NBS) programs. [5] For the purpose of this report, and consistent with previous definitions, NBS "pilot studies" are defined as systematic investigations or public health activities that are designed to evaluate the efficacy and safety of incorporating a new test or condition on a population-based level into state NBS programs.[6]

Obtaining adequate data for newborn population screening is challenging for several reasons. Since most heritable and congenital conditions targeted by NBS are rare, a large number of newborns must be screened to assess the validity and utility of new screening opportunities. Second, most conditions targeted have a substantial degree of clinical variability, creating the need for studies that are sufficiently large to characterize the range of disease variants and false positive presentations that would be found in full-scale population screening programs. Third, pilot studies are usually warranted only after there is some preliminary evidence that early identification followed by treatment or a preventive intervention will provide a net

benefit to the child. At such a point, it may be ethically problematic to conduct a randomized trial of the treatment intervention (that is, randomizing between detection by screening vs. detection through clinical presentation), thus limiting the ability of the pilot study to demonstrate clear utility of screening compared to clinical diagnosis. Fourth, some IRB's reviewing pilot studies in this domain have required the informed consent of parents. A formal or conventional consent requirement reduces uptake into pilot studies, meaning that pilot studies need to be proportionately larger to identify an adequate number of affected infants. Currently, there is no established national system to conduct pilot studies in this domain, resulting in a substantial degree of variability in the design, size, and quality of studies.

For these reasons, the evidence review process used by the ACHDNC is often hampered by limited data on which to make recommendations that will impact the 4 million infants born each year in the United States. To address these challenges, the ACHDNC established a Pilot Studies Workgroup on May 30, 2014. The charge of the Workgroup was to:

- Recognize and support current efforts regarding pilot studies and evaluation
- Identify other resources that could support pilot studies and evaluation
- Identify the information required by the Committee to move a nominated condition into the evidence review process (i.e., define the minimum pilot study data required for a condition to be accepted for evidence review)

This Report represents the report of the Workgroup to the ACHDNC.

II. ACHDNC Evidence Review Process

New conditions are nominated by the lay and/or professional communities for consideration by the ADHDNC for inclusion on the RUSP. The Nomination Package for new conditions includes a form that requires information on the condition, treatment, screening test, confirmatory test, and population-based pilot studies. The Committee's Nomination and Prioritization Workgroup reviews the completed Nomination Package and compiles a summary for Committee consideration. The Committee decides if sufficient evidence is provided to warrant further consideration of the disorder, and votes to assign, or not assign, the nominated condition to the external Condition Review Workgroup. The work of the Pilot Studies Workgroup will assist the ACHDNC in making a determination about whether sufficient data exist from pilot studies to warrant a formal evidence review by the Condition Review Workgroup. If the external Condition Review Workgroup conducts a formal review, it will analyze the available data and create a summary report for ACHDNC. The Committee utilizes the evidence review during deliberations on votes to recommend or not recommend adding the nominated condition to the RUSP for consideration by the Secretary of Health and Human Services. The Public Health Service Act 42 U.S.C. 217a requires that ACHDNC must vote on nominated condition no later than 9 months after having initiated the external evidence review. If the committee then voted to recommend inclusion of the condition into the RUSP, the DHHS Secretary is given 120 days to decide on adoption or rejection of such recommendation.

B. Pilot Studies

Nominations are reviewed for their completeness by staff when submitted. The following describes the outcomes of those nominations that reach the ACHDNC for consideration. The most common reasons that a nominated condition is deemed inadequate to justify external review are:

1) the absence of an adequate pilot study having been done using the same technology as would be used in the US (e.g. spinal muscular atrophy, Fabry disease, and Pompe disease in its 2008 submission);
2) the population on which a pilot was done does not reflect the diversity of the US population (e.g., 2008 Pompe disease pilot studies were done in the Taiwanese population);
3) the length of time over which patients were followed was insufficient to adequately inform outcomes following treatment (e.g., 2008 Pompe disease nomination).
4) the absence of a targeted and specific treatment that significantly impacts patient outcome (e.g., Niemann Pick A/B).

A unique aspect of pilot studies of conditions being considered for NBS is that they often provide the first unbiased ascertainment of patients at risk for genetic disease. Prior to that time, most information about the disease is biased by being from individuals who present to the health care system with disease. Occasionally there are family-based studies that are somewhat less biased in their ascertainment that may raise suspicions of incomplete penetrance or wider disease variation than previously expected. This has been most recently experienced in pilots for lysosomal storage diseases. Incidences of individuals with biochemical evidence of the condition have been higher than previously expected, much of which is adult onset disease forms. This is most pronounced in Fabry disease in which NBS pilots have found it to be much more common than previously expected (1 in 3,000 in Italy; 1 in 1,500 in Asia), with as much as 90% of those being identified in NBS determined to be adult onset forms of disease. When pilot studies provide unanticipated new data about the clinical history of a condition, new questions arise about whether various types of patients can be distinguished and when treatment must be initiated.

III. Types of Data Necessary to Support an Evidence-Review Process

Pilot studies are designed to provide data on one or more of the following elements:

- the clinical characteristics of the condition under consideration,
- the feasibility of high-throughput, high quality laboratory screening,
- the feasibility of conducting population-based screening, diagnosis, and follow-up for the condition, and
- the benefits and harms to the infants screened and their family members.

A. Condition Characteristics

Several features of conditions being assessed for inclusion in NBS programs are critical to an effective evidence review. Many are independent of other parameters of the evidence review while others may be dependent on their relationship to other parameters such as treatment type (e.g., the treatment available that justifies considering NBS is only effective in a subset of the patients with the condition).

Condition-specific features include:

1. The incidence of the condition in the general population and in subpopulations.

2. A goal of NBS is to identify infants before they present in the health care system with symptoms. Hence, understanding the proportion of cases that become clinically affected in the first 48 hours of life are important considerations.

3. Along with disease incidence, the burden of disease in morbidity and mortality provides a sense of the public health importance of screening for a condition.

4. Genetic features of conditions to be considered include the variable expressivity of the disease presentation and the penetrance of the disease in the population.

Data about condition characteristics may come from population-based pilot studies, such as the incidence of the condition, and information about expressivity and penetrance. Other characteristics, such as the burden of disease and age of onset of symptoms, may come from clinical histories.

B. Laboratory Feasibility

Laboratory testing is the central and vital component of newborn screening. Unequivocal demonstration that laboratory testing is feasible in the high-throughput environment of an experienced newborn screening program is a requirement for a successful screening program. The laboratory testing process must be robust and reliable, ensuring accuracy in measurements every day, for every test, on every sample from every baby.

1. **Analytical validation** is a documented process to ensure that a laboratory-developed test is rugged and robust and suitable for its intended use. It assesses and documents test performance characteristics such as Precision, Accuracy, Reportable Range or Linearity, Detection Limit, Interference and Reference Interval. Examples of studies that would address these performance indicators of a biochemical marker are described in Table I. **Analytical Verification** of an unmodified FDA-cleared or FDA-approved test is necessary to verify the performance specifications established by the

vendor. This includes documentation of the following performance characteristics: Precision, Accuracy, Reportable Range and Reference Interval.

Preliminary and Independent Studies that support Performance Characteristics of a Test

a. Preliminary studies for Initial Test Development: Review of published literature may identify tests of biological markers including description of "proof of concept", assessment of optimal test conditions, and identification of key interferences.
b. Evaluation of Methods and Platforms: The feasibility studies provide the opportunity to evaluate adaptations of methods that may exist in a research or diagnostic setting. This process determines which application is likely to work best, and may identify platforms that may be preferable for high birth rate versus low birth rate populations (scalability).
c. Selection of Best High throughput Methods: A technology may perform well in low-volume diagnostic laboratories but may not be sufficiently robust for a high-throughput non-stop environment. Such test performance characteristics are best determined by high-volume laboratories (e.g. public health laboratories) that are capable of translating research tests into robust screening tests during the course of a feasibility analysis.

2. **Clinical Validation** of a previously uncharacterized test within a population is determined within the context of a prospective pilot study. This study investigates the ability of an analytically validated test to detect the clinical condition within the population for which the test measurement is intended (clinical sensitivity) while excluding those within the population who are not affected by that clinical condition (clinical specificity). Useful test performance metrics within the population that are derived from these values include the clinical sensitivity and specificity, and the positive and negative predictive values of the test. The size of the study population is dependent on the incidence of disease within the population. The study requires the availability of confirmatory testing, which will determine true and false positives, and true and false negatives.
    a. **Population Data (Raw Data)**
        i. True positive: patient has the disease and tests positive
        ii. False positive: patient does not have the disease and tests positive
        iii. True negative: patient does not have the disease and tests negative
        iv. False negative: patient has the disease but tests negative

b. **Clinical Performance indicators (Derived Data):**
   i. Clinical Sensitivity refers to the ability of the screening test to correctly identify those patients with disease
      - True positives/(True positives + False negatives)
   ii. Clinical Specificity refers to the ability of the screening test to correctly identify those without disease
      - True negatives/(True positives + False positives)
   iii. Positive Predictive Value addresses how likely a newborn with a screen-positive result actually has the disease
      - True Positives /(True positives + False positives)
   iv. Negative Predictive Value addresses how likely a newborn with a screen-negative test does not have the disease
      - True negatives/(True negatives + False negatives)

3. **Availability of Quality Assurance Materials.** Quality Assurance materials are necessary to ensure continued delivery of high quality testing, and are included during the analytical process to monitor test performance. Feasibility evaluations offer the opportunity for development of new quality assurance testing materials through CDC, for beta-testing the utility of such materials in high performance laboratories and for testing the large scale production of such materials that would be needed for nationwide implementation.

C. Feasibility of a Newborn Screening System

Newborn screening is appropriately referred to as a system due to the various program components that must be integrated that extend from parent and professional education through to diagnosis, care, and follow-up of affected infants. Each target condition requires a somewhat different system due to the nature and variability of the condition, the type and number of false positive results, and the availability of specialists and facilities available for affected infants, among other elements. Further, the NBS system may vary between states due to different ethnic or racial mixes, variations in the availability of specialists and tertiary healthcare facilities for complex conditions, and variations in coverage of treatments by third-party payers.

Pilot studies can provide data on the feasibility of one or more models for a NBS system for a new condition. Demonstration of the feasibility of screening, diagnosis and follow-up in one or a few systems is important, but will not address all of the variables that will be confronted by each state program. Pilot studies should be conducted in populations and using NBS approaches that are sufficiently similar to programs in the US so that valid conclusions can be reached. However, state programs may need to phase in new screening programs in discrete steps in order to demonstrate feasibility within their local environments. (For the purposes of this report, we would not refer to these implementation phases following inclusion on the RUSP as "pilot studies." Other

terms like "implementation studies" or "demonstration projects" might be more appropriate for post-RUSP feasibility studies.)

D. Benefits and Harms

The net clinical benefit from NBS is determined through a comparison of the benefits and harms to newborns from early detection through population screening with the benefits and harms to newborns from clinical detection. In making a recommendation for including a new condition on the RUSP, ACHDNC evaluates both the potential magnitude of the net benefit and the certainty with which the net benefit will be achieved.(3) In general, conditions are approved if the net benefits are predicted to be significant and the certainty is high (assuming feasibility of population screening and adequate readiness of state health programs).

Demonstration of a net benefit from population screening would be best accomplished from an analytic perspective if a randomized controlled trial (RCT) of screening vs. clinical detection is conducted. However, this type of study faces several significant challenges. First, given the clinical variability for many conditions, a comparison of outcomes of an affected population detected through screening to affected population identified through clinical presentation will create a detection bias. In this circumstance, the population detected clinically will be enriched with more severely affected cases because these are the children who will be most readily identified by clinicians. This detection bias can lead to a false conclusion that screening reduces morbidity and mortality because the screened population will be, on average, less severely affected. Therefore, detection bias should be reduced as much as possible in conducting pilot studies of population screening.

A second challenge to RCTs are the ethical concerns about withholding potentially beneficial treatments from newborns randomized to the clinical detection or a later-detection arm of the study. If there are interventions for conditions under consideration that have evidence of efficacy, it would be ethically problematic to withhold the intervention from infants randomized to an unscreened/later-detection group.

There are several options for conducting population-based pilot studies within these constraints. If there is true equipoise regarding the net benefit of early intervention, then a randomized controlled trial may be ethically appropriate. The cystic fibrosis newborn screening pilot study begun in the 1980's in Wisconsin was one of the very few randomized, population-based pilot studies in newborn screening.[7] It used an innovative but controversial design to reduce detection bias by screening all newborns in the state but only looking at the test results for half of the newborns. The other half had their screening results examined at 4 years of age, at which point the affected infants were identified, evaluated, and provided care for CF if they had not been previously diagnosed clinically. This study was instrumental in demonstrating the net benefits of newborn screening for CF.[8]

If a randomized controlled trial is not deemed appropriate, one alternative approach is a comparison of cohorts, one in which screening has been implemented and one that has not undergone screening. Detection bias is a concern with this

design unless measures are taken to identify all affected newborns in each cohort. This approach was used to evaluate newborn screening for neuroblastoma.[9, 10]

A third alternative to an RCT is the use of historical controls as the comparison group. If the clinical variation and the natural history have been adequately characterized for a condition, then outcomes of affected children identified through screening can be compared with historical controls identified clinically.

Finally, a pilot study can be designed to demonstrate feasibility of detection through population-based screening, but not necessarily designed to demonstrate efficacy of early detection. In this context, data on outcomes from early detection from other means are necessary. For heritable conditions, alternative methods of early detection include testing of newborn siblings of affected children or prenatal diagnosis based on a family history. An example is Severe Combined Immunodeficiency Disorder (SCID) for which bone marrow transplantation had been shown to be more effective at younger ages and prior to the onset of systemic infections. A pilot study for SCID was considered sufficient for a decision about inclusion on the RUSP by the ACHDNC after only one newborn had been successfully identified.(11) In this example, the pilot study was deemed successful in demonstrating the feasibility of the NBS screening system for SCID.

IV. Parental Permission for Pilot Studies

Traditionally, many states have retained residual dried bloodspots (DBS) following newborn screening for uses in biomedical research, quality improvement activities, and forensic applications.(12) Research with DBS has been done almost exclusively with de-identified specimens. Under federal research regulations, research with de-identified biospecimens has not been considered human subjects research and therefore has not required the informed consent of parents. This practice has been controversial and led to recent changes in federal law, as described in more detail below. When pilot studies involve identifiable newborns, whether and how parental consent should be obtained has been addressed in various ways.

<u>Parental Permission Approaches</u>

Newborn screening is conducted in almost all states under public health authority and without formal parental permission. However, public health authority does not extend into the practice of research, and as such, the conduct of newborn screening pilot studies raises key questions about whether and how parental permission should be obtained.

Two IRB-approved models for obtaining permissions have been applied with success:

Verbal opt out after written educational materials: The 1980 Wisconsin study of newborn screening for cystic fibrosis was a statewide clinical trial in which infants were randomly assigned to groups where screening results were or were not provided; stopgap protections were in place. New parents were provided information in written form to enable them to opt out of participation.[7]

Verbal consent after written educational material and collector documentation of decision: The 1999 Massachusetts studies of newborn screening for 19 metabolic conditions and for cystic fibrosis, and a later study of newborn screening for SCID, were offered via a statewide research protocol in which nursery staff were required to provide a brochure and to ask a guardian of each infant whether they wanted optional screening(s). Staff were only required to document when a parent said no (due to national level of concerns, Massachusetts expects that staff will be required to document consent in the future).[13].

One IRB-approved model attempted a more conventional research recruitment and consent model and proved to be ineffective for statewide pilot studies:

Individualized consultation with a research coordinator in each participating hospital and guardian signature: The early 2000 California studies of newborn screening for metabolic conditions required multi-site IRB submissions and research coordinators at each site. The model was determined to be unsustainable due to the amount of effort required for recruiting. The California study experienced recruitment rates of only about 60%. While some parents declined participation, a substantial portion of the infants were not included in the study due to hospitals and hospital staff who did not have the time to engage the new parents in a permission process.[14] A recruitment rate of 60% for a rare condition means that studies must be proportionately larger to identify a sufficient number of affected infants to obtain valid results.

This experience illustrates that there are several ways to approach parental permission for pilot studies of new newborn screening modalities. All involve the provision of information to parents. The nature of the decisional process and its documentation differ. Using an "opt-out" approach or a verbal "opt-in" appear to be effective in supporting high recruitment rates. A requirement for a signed consent form has, to date, been associated with reduced recruitment rates.

Recent Changes in Federal Policy*

NBS research is often funded by federal sponsors such as the National Institutes of Health (NIH) and the Health Resources and Services Administration (HRSA). In 2009, parents in two states, Minnesota and Texas, brought suits against their state programs for the state policies and procedures related to secondary uses of residual newborn screening bloodspots.[17] Residual dried bloodspots (DBS) are present for almost all infants screened and state programs manage these biospecimens in different ways. Potential uses of residual DBS include biomedical research, QA/QI applications, and forensic uses.[1] For states that retain DBS, this was traditionally done without specific parental permission because NBS is conducted without parental permission. Many states that retain DBS for secondary purposes will notify parents of this practice in their informational brochures but many parents do not read or fully understand this information. Further, the federal human subjects regulations permit biospecimens to be used without the permission

of the tissue source, or their family members, if the specimens are de-identified or if the research meets the criteria for waiver of informed consent.[45CFR46] Most research in this domain is conducted with de-identified specimens. Research to assess public attitudes suggest that most people want to be informed about their state policies and practices and indicate that parents should have a choice.

With the renewal of the Newborn Screening Saves Lives Reauthorization Act in 2014, an amendment was included that specifically requires the informed consent of parents for research uses of newborn screening DBS that are funded by the federal government.[18] Permission is required regardless of whether the DBS are de-identified or if the research meets the criteria for a waiver of consent. These new requirements went into effect in March of 2015, and they prohibit the use of all DBS collected after that in federally funded research. Given the lack of policies and procedures for informed consent for DBS use in most states, DBS collected after the implementation of the new law will not be available for NBS research until a permission process is in place. At the time of this writing, the federal Office of Human Research Protection (OHRP) is drafting guidelines for parental permission for research uses of DBS. While parental permission is supported by the public and required by federal law, this requirement presents a substantial challenge for NBS programs to conduct new pilot studies using DBS.

**45CFR46 has been revised since the development of this report. Information was up to date as of May 2016.*

V. Current Activities at the Federal Level that Support Pilot Studies

A. NIH

The primary focus of Newborn Screening at the NIH is guided by the language of the Newborn Screening Saves Lives Act. The language encourages the NIH, under the auspice of the Hunter Kelly Newborn Screening Research Program, to continue carrying out, coordinating, and expanding research in newborn screening in the following areas:

A. identifying, developing, and testing the most promising new screening technologies, in order to improve already existing screening tests, increase the specificity of newborn screening, and expand the number of conditions for which screening tests are available;

B. experimental treatments and disease management strategies for additional newborn conditions, and other genetic, metabolic, hormonal, or functional conditions that can be detected through newborn screening for which treatment is not yet available; and

C. providing research findings and data for newborn conditions under review by the Advisory Committee on Heritable Disorders in Newborns and Children to be added to the recommended uniform screening panel;

D. conducting pilot studies on conditions recommended by the Advisory Committee on Heritable Disorders in Newborns and Children to ensure that screenings are ready for nationwide implementation.

Currently, multiple institutes across the NIH support research that addresses components necessary for a condition to be nominated to the Recommended Uniform Screening Panel (RUSP), including but not limited to the identification of disorders and their underlying mechanisms, the development of new and novel screening tests and treatments, and research exploring the ethical, legal and social implications related to newborn screening.

Currently funded activities at the NIH related to pilot studies include:

The Newborn Screening Translational Research Network (NBSTRN)

The Newborn Screening Translational Research Network (NBSTRN) is a resource for investigators engaged in newborn screening-related research, and serves as a comprehensive research infrastructure to facilitate the translation of the research to the newborn screening community. The NBSTRN Coordinating Center is a key component of the the Hunter Kelly Newborn Screening Research Program. Currently, the NBSTRN provides an array of tools and resources including the Virtual Repository of Dried Blood Spots (VRDBS), the Longitudinal Pediatric Data Resource (LPDR), the Laboratory Performance Database, and support for Ethical, Legal and Social Implications of newborn screening research. The LPDR has also developed standardized data elements for over 80 disorders, including all disorders that are part of the RUSP and those in consideration for addition to the RUSP. Experts and public constituents in specific disease groups have participated in the development of these data elements. In collaboration with the National Library of Medicine the data elements have been mapped using existing registries, FDA requirements, and other resources. The data elements and case report forms are publicly available through NBSTRN.

Pilot Studies for Newborn Screening

NICHD is supporting an IDIQ (indefinite delivery/indefinite quantity) master task order to maintain a pool of high-throughput newborn screening laboratories with the capacity to screen, in relatively short periods of time (12-18 months), a large number of newborns (at least 50,000) that are representative of various regions of the United States. Three states (Georgia, North Carolina, and Massachusetts) have been selected to operate under a contract to rapidly develop protocols and initiate testing shortly after the addition of a new condition to the RUSP, with the goal of being able to provide part of the evidence base for addition of high-priority conditions to the RUSP . The contract is currently funding pilots for X-linked Adrenoleukodystrophy (X-ALD) and Spinal Muscular Atrophy (SMA).

Newborn Sequencing In Genomic medicine and public HealTh (NSIGHT)

This pilot program aims to explore, in a limited but deliberate manner, the implications, challenges and opportunities associated with the possible use of

genomic sequence information in the newborn period. The four funded projects are exploring:

- Acquisition and analysis of genomic datasets that expand considerably the scale of data available for analysis in the newborn period.
- Clinical research that will advance understanding of specific disorders identifiable via newborn screening through promising new DNA-based analysis.
- Research related to the ethical, legal and social implications (ELSI) of the possible implementation of genomic sequencing of newborns.

The pilots are expected to provide evidence to answer at least one of the following questions:

- For disorders currently screened for in newborns, how can genomic sequencing replicate or augment known newborn screening results?
- What knowledge about conditions not currently screened for in newborns could genomic sequencing of newborns provide?
- What additional clinical information could be learned from genomic sequencing relevant to the clinical care of newborns?

Active initiatives currently exist at NIH to fund research related to components necessary for addition of new conditions to the RUSP.

Natural History of Disorders Identifiable by Screening of Newborns (R01) PAR-16-061, PAR-18-090

This initiative encourages applications that propose to develop studies that will lead to a broad understanding of the natural history of disorders that already do or could potentially benefit from early identification by newborn screening.

Innovative Therapies and Tools for Screenable Disorders in Newborns (R01, R03, R21) PAR-18-689, PAR-18-690, PAR-18-691

This initiative encourages proposed research relevant to the basic understanding and development of novel screening approaches and therapeutic interventions for currently screened conditions on the RUSP and "high priority" genetic conditions for which screening could be possible in the near future.

NIH Parent Announcement

The majority of grants at NIH are investigator-initiated, and there are many projects related to Newborn Screening funded through this route.

B. CDC

The CDC supports public health laboratories in providing critical laboratory services, technical training and quality materials to ensure rapid and accurate implementation of screening for new conditions, and has also provided funding for implementation of new programs.

CDC can use existing relationships and demonstrated expertise to:

- Help state programs obtain appropriate laboratory equipment and staff infrastructure, allowing states to implement accurate testing for new conditions faster
- Assist states with test validation by creating innovative disease-specific "Validation Test Packages" that can be used by state labs to determine performance characteristics that would ensure delivery of consistent, high quality results
- Create enhanced quality assurance materials that reflect the increasing complexity of disease targets (such as molecular and enzymatic biomarkers) to ensure state newborn screening tests for new conditions are accurate
- Create Newborn Screening "Laboratories of Excellence" that will partner with CDC to help resolve the unique challenges associated with screening for specific diseases and serve as models for other state programs as they implement testing. These early adopting programs will provide additional insight for later adopting programs about implementation of new conditions, which will include algorithms for testing, multiplexing assays, second and third tier testing, educational strategies and short term follow-up.

C. FDA

FDA has a long history of working collaboratively with manufacturers on newborn screening devices and the types of studies needed to support the analytical and clinical validation of those tests. Further, FDA is supportive of studies evaluating investigational devices, including pilot studies. (Investigational devices are test systems that have not established their analytical and/or clinical validity for its proposed use.) FDA encourages manufacturers and laboratories that are developing new tests to get early feedback on their pilot studies through the pre-submission process. [15]

There are some types of pilot studies that may need an Investigational Device Exemption [21 CFR part 812] (IDE) from the FDA before the study begins (e.g. if the clinical investigation uses a significant risk device.[15][16] Certain studies that are not "for a use of substantial importance in diagnosing, curing, mitigating, or treating disease, or otherwise preventing impairment of human health and which does not present a potential for serious risk to the health, safety, or welfare of a subject" (21 CFR 812.3(m)(3)) do not need FDA approval to begin. Certain pilot studies of investigational devices in newborn screening have not needed an FDA approved IDE but still had to follow the abbreviated requirements (including informed consent) listed at 21 CFR 812.2(b)).

VI. Recommendations

Charge 1) Identify the information required by the Committee to move a nominated condition into the evidence review process (i.e., define the minimum pilot study data required for a condition to be accepted for evidence review).

This charge clearly articulates that our task is to define a threshold of pilot study data to move a condition to a formal evidence review. Our task is not to define

what data must exist to warrant a positive recommendation for inclusion of a new condition on the RUSP.

As noted, we are defining "pilot studies" as those studies, whether conducted under a research paradigm or under a public health paradigm, that are conducted prior to broad-scale, population-based screening under the authority of state public health service newborn screening programs. Therefore, pilot studies include evaluations of testing modalities and population screening approaches. The evaluation of population screening approaches must be designed to demonstrate feasibility of screening at this volume and not necessarily demonstrate the efficacy of post-screening interventions if the interventions have been otherwise evaluated for net benefit.

Recommendation 1) Data should be available on the analytical validation of one or more screening modalities proposed for use in population-based screening in newborns. Data should include information on precision, accuracy, the reportable range, detection limits, interference, reference intervals, and cost. Pilot studies for analytical validation should include use of dried bloodspots, or other biological specimens or physiologic assessments, from a population of newborns (or other target populations), including known true positive and negative specimens, in addition to laboratory prepared target specimens.

Recommendation 2) Data should be available on the net benefits of clinical interventions following early detection compared to clinical diagnosis. Early detection can be achieved through population screening pilot studies, through testing secondary to a family history of the condition, or through targeted screening of high-risk groups.

Recommendation 3) Data should be available from pilot studies involving population-based screening of identifiable newborns.

3A) The study should be sufficiently large to identify at least one true positive, clinically affected newborn for the condition under consideration, and

3B) The population included in the pilot study, and the screening protocol used, should be similar to the US population and to state NBS programs with respect to known prevalence of the condition, and the timing and approach to screening. The screening modality used in the pilot study or studies should be comparable to the method proposed in the application.

Charge 2: To recognize and support current efforts regarding pilot studies and evaluation

As noted in this report, extensive efforts at the NIH, CDC, and FDA are being conducted that support newborn screening pilot studies.

Recommendation 4) Continued support should be provided by DHHS for the NIH initiatives relevant to pilot studies in newborn screening including the

NBSTRN, NSIGHT, the Pilot Studies grants, Natural History grants, Innovative Therapies grants, and grants supported under the Parent Announcement.

Recommendation 5) Continued support should be provided by DHHS to the CDC for its activities relevant to pilot studies that address technical training and quality materials for state laboratories, assistance to state and other programs in obtaining laboratory equipment, the creation and distribution of "Validation Test Packages," population surveillance, and the fostering of "Laboratories of Excellence."

Charge 3) Identify other resources that could support pilot studies and evaluation. This Report documents the existence of an extensive set of resources available for the conduct of pilot studies in newborn screening at the federal and state levels. The critical element lacking in the current environment is a stable infrastructure for pilot studies that could involve a coordinated network of state programs and professionals, including laboratories, clinical experts, public health professionals, research methodologists, bioethicists, lay advocates, and Institutional Review Boards (IRBs).

Components of a national infrastructure currently exist through the work of the NBSTRN, NIH's support for pilot studies, and CDC's work to support test evaluation and quality assurance. Establishment of a coordinated network will minimize delays and uncertainties created by "re-inventing the wheel" with the design, coordination, and conduct of each large pilot study.

Recommendation 6) DHHS should support the development of a research network comprised of state-based public health programs, laboratories, and academic or other research centers that would provide a stable, experienced, compliant, efficient, and quality infrastructure for the conduct of population-based pilot studies for newborn screening.

Each "Center of Excellence for Newborn Screening Pilot Studies" would consist of local or regional collaboration between a newborn screening program, a laboratory, and a research center. When a new condition is a candidate for future inclusion on the RUSP, the Centers of Excellence could collaborate to develop and conduct the pilot studies to provide the necessary data for an evidence review. Such a system could overcome many of the persistent challenges in the conduct of pilot studies by enabling the recruitment of sufficient number of participants to address rare conditions, the provision of expertise and experience in clinical medicine and relevant research methodologies, the use of a consistent approach to parental education and permission, the receipt of input from advisory groups with relevant stakeholders, the provision of expertise and services from experienced newborn screening laboratories, and the involvement of IRBs that are experienced with the review and oversight of such pilot studies. Centers located in states with large, diverse populations could conduct pilot studies with a single protocol or could be designed to compare screening approaches, as appropriate. Such a system could

provide high-quality data in an efficient fashion to support robust evidence-review processes at the ACHDNC and for state newborn screening programs.

References

1) American Academy of Pediatrics. Serving the family from birth to the medical home. Newborn screening: a blueprint for the future—a call for a national agenda on state newborn screening programs. Pediatrics. 2000;106(2 pt 2):389–422
2) http://www.hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/about/policiesprocedures.pdf
3) Perrin, J. M., Knapp, A. A., Browning, M. F., et al. (2010). An evidence development process for newborn screening*., 12*(3), 131-134.
4) Goldenberg AJ, Comeau AM, Grosse SD, Tanksley S, Prosser LA Ojodu J, Botkin JR, Kemper AR, Green NS. Evaluating Harms in the Assessment of Net Benefit: A Framework for Newborn Screening Condition Review. Matern Child Health J. 2016 Mar;20(3):693-700. doi: 10.1007/s10995-015-1869-9.
5) Pass K, Green NS, Lorey F, Sherwin J, Comeau AM. Pilot studies in newborn screening. Ment Retard Dev Disabil Res Rev. 2006;12(4):293-300.
6) Botkin JR, Lewis MH, Watson MS, Swoboda KJ, Anderson R, Berry SA, Bonhomme N, Brosco JP, Comeau AM, Goldenberg A, Goldman E, Therrell B, Levy-Fisch J, Tarini B, Wilfond B. Parentall permission for pilot newborn screening research: Guidelines from the NBSTRN. Pediatrics 2014;133:e410–e417
7) Taylor HA, Wilfond BS. Ethical issues in newborn screening research: lessons from the Wisconsin cystic fibrosis trial. J Pediatr. 2004;145(3):292–296.
8) Gross SD, Boyle CA, Botkin JR, Comeau AM, Kharrazi M, Rosenfeld M, Wilfond BS. Newborn screening for cystic fibrosis: Evaluation of benefits and risks and recommendations for state newborn screening programs. MMWR 2004;53(RR13): 1-36.
9) Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. N Engl J Med. 2002;346:1047–1053 34.
10) Woods WG, Gao RN, Shuster JJ, et al. Screening of infants and mortality due to neuroblastoma. N Engl J Med. 2002;346:1041–1046.
11) Hale JE, Bonila FA, Pai SY, Thompson JIG, Notaraneglo LD, Eaton RB, Comeau AM. Identification of an infant with severe combined immunodeficiency by newborn screening. J Allergy Clin Immunol 2010; 125(5):1073-1074.
12) Lewis M, Goldenberg A, Anderson R, Rothwell E, Botkin J. State Laws Regarding the Retention and Use of Residual Newborn Screening Blood Samples. Pediatrics 2011;127:703-712.
13) Comeau AM, Levin D. Population-based research within a public health system: two models for common rule compliance in the Massachusetts Newborn Screening Program. In: Bailey M, Murray T, eds. Ethics and Newborn Genetics Screening: New Technologies, New Challenges. Baltimore, MD: Johns Hopkins Press; 2009:274–291
14) Feuchtbaum L, Cunningham G, Sciortino S. Questioning the need for informed consent: a case study of California's experiencewith a pilot newborn screening research project. J Empir Res Hum Res Ethics. 2007;2(3):3–14
15) http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM311176.pdf
16) http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126418.pdf

17) Lewis MH. Lessons from the residual newborn screening dried blood sample litigation. J Law Med Ethics 2015;43 (Suppl 1):32-5.
18) https://www.congress.gov/bill/113th-congress/house-bill/1281/text

Acknowledgements

The Pilot Studies Workgroup would like to thank Joseph Bocchini, Debi Sarkar, Alaina Harris, Melissa Parisi, and Andrea Atherton for their contributions to this report.

Appendix A: Analytical Validation of a Screening Test Documented process to ensure that a method is robust and is suitable for its intended use.	
Precision	**AIM:** Determine the agreement of replicate measurements and assess the amount of "imprecision" or random error in the method. **Approach:** Create 20 samples that cover the measurement range of the test. "Determine Within day variation": Test these samples by running repeatedly (20 times) within the period of one day. "Determine Day-to-day variation": Test samples by running them every day for 20 days.
Accuracy	**AIM:** Determine how close the measurements are to the true value and assess "inaccuracy" or systematic error in the method. **Approach:** Test different 20 patient specimens by using your new method and a "reference method". Repeat this same analysis for 20 days. **Alternate approach if there is no reference method.** **AIM:** Perform Recovery Study. **Approach:** Spike known quantities of marker into known volume of blood and create DBS. Perform test to determine the amount of sample you detect and compare with expected values.
Reportable range or linearity	**AIM:** Establish the reliability of the highest and lowest test results. **Approach:** Create 2 pools of purchased blood, one at baseline levels of the marker and one spiked with a high concentration of the marker. Make dilutions between these two initial pools to cover the possible range of measurements made by the test. Spot blood on filter paper and evaluate the highest and lowest concentration that can be reliably be determined.
Detection Limit	**AIM:** Determine the lowest concentration of the marker that can be measured using conditions of your test. Also called the Analytical sensitivity. **Approach:** Create a blood pool that has been spiked with the marker of interest and sequentially dilute the sample. Create DBS using these samples and test to determine how reliably you can detect these samples.
Interference	**AIM:** All patient samples are not "perfect" and the presence of "interferences" can result in inaccuracies in measurement. This addresses Analytical Specificity. **Approach:** Create DBS samples that contain common interferences (such as bilirubin, hemolysis, lipemia). Test samples to determine the effect of these known interferences on the accuracy of your measurement.
Reference interval	**AIM:** Reference Interval study should reflect the laboratory's testing population. It is established for each population to determine the distribution of measurements for the normal population, compared with values in the diseased population. **Approach:** Test several thousand samples to determine the range of measurements within that particular population. Include the following sample types: (1) normal samples; (2) samples from known patients and laboratory-created affected samples; (3) samples from patients with related diseases. Reference intervals are dependent on several factors such as choice of testing platform, test conditions (time, temperature, reagent concentrations), age at collection, choice of standard curve or calibrators.

Appendix B

Special considerations for laboratory testing that can impact implementation of population testing.

- Access to Reagents and Cost of Test**:** Feasibility studies provide the opportunity to understand and prevent potential bottlenecks in reagent production, access to quality control materials, access to quality assurance materials and access to or performance of instruments required for screening. The cost of the test in a high throughput setting can also be assessed.
- Algorithm Development: Feasibility studies offer the opportunity to test the sequence of operations prior to the issuing of a report. For example, does the laboratory testing process require (for some or all of the samples): (a) retesting of the same specimen by the same assay, (b) retesting of the same specimen by an alternative assay (second and third-tier tests).
- Integration to Laboratory Information Management System (LIMS) system: Although generalized planning for LIMS can begin, until the laboratory testing and follow up module is defined, programming cannot go forward. Once a strong foundation for the laboratory testing algorithm is developed within the feasibility study, programming can be more fully developed and validation of programming rules can be optimized during the feasibility study period. In situations where the testing algorithm is similar to already-existing algorithms, the process is less complicated.
- Infrastructure and Special Issues: The implementation of each new condition raises new unexpected issues which need to be addressed within the context of population based screening. The Public Health setting has ensured management of communications between healthcare providers who are unlikely to be expecting to have to act on laboratory results. Unlike the diagnostic setting, providers typically do not order the tests and typically are seeing asymptomatic neonates. This context alone brings forward a variety of issues for testing, communications and follow up that have been efficiently and successfully addressed by the experts working within state newborn screening programs.
- Case Definition and Incidental Findings: Diseases identified through newborn screening often reveal an unexpected spectrum of clinical presentations (only identified once screening is initiated). The specific condition that is the focus of the screen must have a pre-defined case definition in order to determine the efficacy of the screen. Feasibility studies may help uncover the spectrum of disease that may be observed by screening and help define the screen's target condition.
- Testing Availability for diagnostic follow-up: Feasibility evaluations provide indications of availability and quality of diagnostic evaluations. The evaluation will show whether diagnostic centers can handle the volume of new patients who must be evaluated after having received a positive result on a newborn screen. Key to this evaluation is the question of whether the diagnostic evaluation can be reliably performed on a newborn and whether the diagnostic

criteria for meeting the case definition have been established for asymptomatic newborns.

Appendix C

Examples of ACHDNC Reviews and the Role of Pilot Studies

Mucopolysaccaridosis Type I (MPS I),

MPS I is an autosomal recessive Lysosomal Storage Disorder, and a progressive condition affecting multiple organ systems. In its most severe form, MPS I often results in death or severe disability. Insufficient enzyme activity prevents the proper recycling process, resulting in the storage of materials in virtually every cell of the body. Enzyme replacement therapy (ERT) and hematopoietic stem cell transplantation (HSCT) are the standard modes of treatment.

In February 2015, the ACHDNC's Evidence Review Group presented data from five separate pilot studies from Missouri (149,500 total screened), Illinois (17,300), Washington (106,526), Taiwan (35,285) and Italy (3,403). These studies were critical to the committee's deliberations because they demonstrated the feasibility of several different screening methods (e.g., digital microfluidics, MS/MS, fluorescence assay), demonstrated the likelihood of positive screens and other outcomes (confirmed MPS I, carriers, false positives, pseudodeficiencies), helped refine screening algorithms to improve accuracy of screening, and demonstrated that although screening is technically feasible, it will be a challenge for states to implement in the near future. The committee recommended that MPS I be added to the RUSP, but also recognized the need for additional pilot studies and technical assistance so that states could further define the most appropriate test platform and laboratory protocol, and establish short and long term follow up procedures.

X-Linked Adrenoleukodystrophy (X-ALD) X-ALD is a peroxisomal disorder that affects the adrenal cortex and the central nervous system. X-ALD has a broad phenotypic spectrum, but males with childhood cerebral ALD (CALD) have extensive myelopathy and severe behavioral and cognitive disorders. Untreated CALD typically leads to death within three years of onset. The primary treatment for X-ALD is hematopoietic stem cell transplantation. Adrenal cortisol replacement therapy is also necessary for adrenal insufficiency.

In 2012 the ACHDNC reviewed the initial nomination for X-ALD but did not send it forward for evidence review, citing the lack of pilot study data. The evidence review that was completed in 2015 reported data from the New York state newborn screening program. A total of 363,755 newborns were screened during an 18-month period. The pilot study demonstrated the usefulness of a "3-tier" screening program to reduce false positives. The pilot study ultimately found 33 screen-positive children. Of those 14 (42%) were males with ABCD1 mutations, 14 (42%) were female carriers of an ABCD1 mutation, 3 (9%) had a Zellweger Spectrum

disorder, and 1 (3%) had Aicardi-Goutieres syndrome (another condition associated with leukodystrophy). X-ALD was considered again by ACHDNC and recommended for the RUSP in September 2015.

SCID

Severe Combined Immunodeficiency (SCID) was one of two conditions put forward for nomination in 2007, just as the Evidence Review Workgroup (ERW) for the SACHDNC and the SACHDNC itself were beginning to develop the first formal processes for the evaluation of conditions nominated for newborn screening. The nomination of SCID was formally accepted and sent to the ERW in January, 2008. At the February 2009 meeting, the SACHDNC recognized the following major gaps in evidence that had been noted in the final report submitted by the ERW: [*http://www.hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/nominatecondition/reviews/severeimmunodeficiencyreport.pdf*].

- Absence of a confirmed case of SCID identified through a population-based newborn screening pilot study.
- Absence of a documented willingness and capacity to implement SCID newborn screening by states other than the two early adopters.
- Absence of documented reproducibility of assay performance data.
- Absence of proficiency testing samples available for Quality Assurance Monitoring.

Using the then-recently approved processes adopted by the SACHDNC [*Perrin JM;Knapp AA; Browning MF et al. An evidence development process for newborn screening. Genetics in Medicine Vol 12 Number 3 March 2010 pp 131-134; Calonge N, Green NS, et al. Committee Report: Method for evaluating conditions nominated for population-based screening of newborns and children. Genetics in Medicine Volume 12 Number 3, pp 153-159 March 2010*], and the simple decision matrix outlined in Table 1 of the latter, the SACHDNC determined that the evidence was insufficient to make a recommendation to add SCID to the core panel, but was compelling enough to recommend additional studies to fill in the evidence gaps.

At the January 2010 meeting, the SACHDNC accepted a revised nomination for recommending that both SCID and related T-cell lymphopenias be put on the "core" and "secondary panel" of the RUSP, respectively. Population-based pilot studies continued in Wisconsin and Massachusetts. The SACHDNC reasoned that since both states had documented the identification of infants who had T-cell lymphopenia through newborn screening, the required demonstration that a confirmed case would be identified by screening had been provided. Massachusetts reported it had provided a week-long training session for the newborn screening programs of Texas, California and Minnesota, documenting other states' willingness and capacity to implement SCID newborn screening. The combined data from Wisconsin and Massachusetts screening experiences (denominator of screened infants, number of screened infants with a)findings prompting re-test of the same sample, b)findings prompting request for new sample, c) findings prompting diagnostic evaluation by

flow cytometry, d) findings prompting further diagnostic evaluation, e) findings yielding a SCID case or separately, a TCL case) demonstrated reproducible assay and algorithm performance data and CDC reported their readiness to provide proficiency testing materials. The SACHDNC letter to recommend that SCID and TCL be added to the RUSP, was sent to HHS Secretary Sebelius Feb 25,2010.

Krabbe Disease

Krabbe disease (OMIM #245200) is a devastating autosomal recessive, demyelinating condition. In its classic form, patients present in the first few months of life with irritability, spasticity, and progressive motor and mental deterioration causing death in childhood. In addition to this early presentation, later to adult onset and milder variants exist. The diagnosis is traditionally based on the finding of deficient lysosomal galactocerebrosidase (GALC) activity in leukocytes in a patient with a suspicious phenotype. Treatment is currently limited to supportive care but hematopoietic stem cell transplantation (HSCT), especially when performed early in life, has some benefit. In 2006, after a high-throughput GALC enzyme assay in DBS became available, Krabbe disease was included in the New York state (NY) screening program. In 2008, the ACHDNC initiated an evidence review that was largely based on the findings from the NY screening program. At that time, nearly 770,000 newborns had been screened using a two-tier approach that included the GALC enzyme assay as the primary test followed by molecular genetic analysis of the GALC gene when GALC activity was below the chosen cutoff. 236 cases (0.03%) required molecular genetic testing. Of these 140 (0.02%) were referred for follow up because they carried at least one mutation, many that were novel (variants of uncertain significance). Following clinical and laboratory investigations, 7 infants were considered at high, 13 at moderate and 36 at low risk of developing Krabbe disease. Contrary to previous prevalence expectations of ca. 1:100,000 live births and a preponderance of early infantile vs. later onset disease variants (10:1), of the high risk patients only 2 were diagnosed with early infantile Krabbe disease. Both underwent HSCT which one patient did not survive. Another patient left the US, 3 had remained asymptomatic during follow up and one patient refused further investigations. All at-risk cases were to be followed according to guidelines developed by a NY consortium of specialists in the diagnosis and treatment of Krabbe disease. These guidelines include regular clinical, laboratory, neurophysiological and radiologic studies to determine if and when HSCT should be recommended. Upon consideration of the evidence review, the ACHDNC in 2009 rejected the addition of Krabbe disease to the RUSP because it found that sufficient evidence was lacking to support net benefit of both screening and treatment (http://www.hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/nominatecondition/workgroup.html).

Nevertheless, screening continued in NY and through state based legislative action screening for Krabbe disease has been mandated and already commenced in several newborn screening programs. Refinements of the newborn screening approach and follow up process have since been proposed and have been implemented in at least one newborn program (KY) *(Turgeon CT, et al. Measurement of psychosine in dried blood spots--a possible improvement to newborn screening programs for Krabbe*

disease. J Inherit Metab Dis. 2015; 38: 923-9). Concerns regarding the effectiveness of current treatment options remain with early gene therapy approaches potentially offering an improvement in the future *(Ungari S, et al. Design of a regulated lentiviral vector for hematopoietic stem cell gene therapy of globoid cell leukodystrophy. Mol Ther Methods Clin Dev. 2015; 2: 15038)*.