Overview of Cutoff Determinations and Risk Assessment Methods used in Dried Blood Spot Newborn Screening

This publication was supported by Cooperative Agreement # 5NU60OE000103 funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC or the Department of Health and Human Services.
### Contents

Acknowledgements ................................................................................................................................. 2  
Acronyms .................................................................................................................................................. 3  
1. Purpose ............................................................................................................................................... 4  
   1.1 Limitations .................................................................................................................................... 4  
2. Background ......................................................................................................................................... 4  
   2.1 Cutoffs ......................................................................................................................................... 5  
3. Overview of Cutoff Determination ...................................................................................................... 7  
   3.1 Testing Dried Blood Spots from Newborns .................................................................................. 7  
   3.2 Analyzing Data and Determining a Preliminary Cutoff .............................................................. 7  
   3.3 Validating or Verifying the Cutoff ............................................................................................. 8  
   3.4 Special Considerations ................................................................................................................. 9  
   3.5 Collaborative Laboratory Integrated Reports ........................................................................... 11  
4. Cutoffs for Specific Newborn Screening Disorder Categories .......................................................... 13  
   4.1 Endocrine Disorders (Congenital Hypothyroidism, Congenital Adrenal Hyperplasia).............. 13  
   4.2 Cystic Fibrosis ............................................................................................................................ 13  
   4.3 Hemoglobinopathies .................................................................................................................... 14  
   4.4 Enzyme Deficiency Disorders ...................................................................................................... 14  
   4.5 Amino Acid, Fatty Acid Oxidation and Organic Acid disorders ............................................. 15  
   4.6 Severe Combined Immunodeficiency (SCID) .......................................................................... 15  
   4.7 X-Linked Adrenoleukodystrophy (X-ALD) ............................................................................ 16  
5. Monitoring and Ongoing Evaluation of Cutoffs ................................................................................. 17  
6. References .......................................................................................................................................... 18
Acknowledgements

APHL would like to recognize the following individuals that generously contributed to the development of this document and the members of the APHL Newborn Screening and Genetics in Public Health Committee and the APHL Newborn Screening Quality Assurance and Quality Control Subcommittee for their time and input.

Mei Baker, MD, FACMG Wisconsin State Laboratory of Hygiene
Joanne Bartkus, PhD, Minnesota Department of Health
Christianne Biggs, MS, Oregon State Public Health Laboratory
Natasha Bonhomme, Chief Strategy Officer, Genetic Alliance
Michele Caggana, ScD, FACMG, New York State Department of Health- Wadsworth Center
Neena Champaigne, MD, Greenwood Genetics Center
Bob Currier, PhD, California Department of Public Health
Carla Cuthbert, PhD, FCMCG, FACMG, Newborn Screening Molecular Biology Branch, CDC
Tim Davis, Washington State Department of Health
George Dizikes, PhD, HCLD/CC (ABB), Tennessee Department of Health
Christine Dorley, MSP, MT (ASCP) Tennessee Department of Health
Amy Gaviglio, MS, CGC, Minnesota Department of Health
Fizza Gulamali-Majid, PhD, Maryland Department of Health and Mental Hygiene
Patrice Held, PhD, Wisconsin State Laboratory of Hygiene
Amy Hietala, MS, Minnesota Department of Health
Patrick Hopkins, Missouri Public Health Laboratory
Patricia Hunt, Texas Department of State Health Services
Rachel Lee, PhD, Texas Department of State Health Services
Adrienne Manning, Connecticut Department of Public Health Laboratory
Joanne Mei, PhD, Newborn Screening Molecular Biology Branch, CDC
Darren Michael, PhD, Colorado Department of Public Health and Environment
Christine Moore, Texas Department of State Health Services
Hari Patel, MS, North Carolina Department of Health and Human Services
Joseph Orsini, PhD, New York State Department of Health- Wadsworth Center
Kostas Petritis, PhD, Newborn Screening Molecular Biology Branch, CDC
Michael Ramirez, State Hygienic Laboratory at The University of Iowa
Debi Sarkar, MPH, Division of Services for Children with Special Healthcare Needs, HRSA
Inderneel Sahai, MD, University of Massachusetts Medical School
Santosh Shaunak, Washington State Department of Health
Eleanor Stanley, MT (ASCP), Michigan Department of Health and Human Services
Bonita Taffe, PhD, Florida Department of Health
# Acronyms

The following is a list of acronyms used throughout this document:

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino Acid</td>
</tr>
<tr>
<td>AC</td>
<td>Acylcarnitine</td>
</tr>
<tr>
<td>APHL</td>
<td>Association of Public Health Laboratories</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>CAH</td>
<td>Congenital Adrenal Hyperplasia</td>
</tr>
<tr>
<td>CCHD</td>
<td>Critical Congenital Heart Disease</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CH</td>
<td>Congenital Hypothyroidism</td>
</tr>
<tr>
<td>CF</td>
<td>Cystic Fibrosis</td>
</tr>
<tr>
<td>CLIA</td>
<td>Clinical Laboratory Improvement Amendments</td>
</tr>
<tr>
<td>CLIR</td>
<td>Collaborative Laboratory Integrated Reports</td>
</tr>
<tr>
<td>CLSI</td>
<td>Clinical and Laboratory Standards Institute</td>
</tr>
<tr>
<td>Cq</td>
<td>Cycle of quantitation</td>
</tr>
<tr>
<td>DBS</td>
<td>Dried Blood Spot</td>
</tr>
<tr>
<td>FIA</td>
<td>Flow Injection Analysis</td>
</tr>
<tr>
<td>Hb</td>
<td>Hemoglobinopathies</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>IEF</td>
<td>Isoelectric Focusing</td>
</tr>
<tr>
<td>IRT</td>
<td>Immunoreactive Trypsinogen</td>
</tr>
<tr>
<td>LIMS</td>
<td>Laboratory Information Management System</td>
</tr>
<tr>
<td>LPCs</td>
<td>Lysophosphatidylcholines</td>
</tr>
<tr>
<td>LSDs</td>
<td>Lysosomal Storage Disorders</td>
</tr>
<tr>
<td>MoM</td>
<td>Multiples of Median</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem Mass Spectrometry</td>
</tr>
<tr>
<td>NBS</td>
<td>Newborn Screening</td>
</tr>
<tr>
<td>NewSTEPs</td>
<td>Newborn Screening Technical assistance and Evaluation Program</td>
</tr>
<tr>
<td>NICU</td>
<td>Neonatal Intensive Care Unit</td>
</tr>
<tr>
<td>NSQAP</td>
<td>Newborn Screening Quality Assurance Program</td>
</tr>
<tr>
<td>QA</td>
<td>Quality Assurance</td>
</tr>
<tr>
<td>QC</td>
<td>Quality Control</td>
</tr>
<tr>
<td>R4S</td>
<td>Region 4 Stork Laboratory Performance Database</td>
</tr>
<tr>
<td>RUSP</td>
<td>Recommended Uniform Screening Panel</td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis Software</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe Combined Immunodeficiency</td>
</tr>
<tr>
<td>SOPs</td>
<td>Standard Operating Procedures</td>
</tr>
<tr>
<td>TPN</td>
<td>Total Parenteral Nutrition</td>
</tr>
<tr>
<td>TREC</td>
<td>T-cell receptor excision circles</td>
</tr>
<tr>
<td>X-ALD</td>
<td>X-Linked Adrenoleukodystrophy</td>
</tr>
<tr>
<td>µL</td>
<td>Microliter</td>
</tr>
</tbody>
</table>
1. Purpose

The primary purpose of this document is to provide a general overview of the variety of approaches utilized by newborn screening (NBS) laboratories and program staff to determine if a newborn is at risk for a screened disorder. An overview of historical and current approaches that laboratories rely on for risk assessment are described as well as factors that should be considered when establishing and evaluating risk. This document is not intended to provide detailed instructions for performing risk assessment in newborn screening.

1.1 Limitations

Newborn screening is performed as a means to assess risk in clinically asymptomatic infants for a host of disorders\(^1\); it is not intended to establish diagnosis. Abnormal biomarker levels, identified through screening and evaluated using cutoffs, only indicate that a newborn may be at higher risk for a screened disorder. This information is communicated to the healthcare provider, who may request additional diagnostic testing to determine if the newborn has the disorder in question. Healthcare providers should understand that in instances where a newborn has a family history of a disease or is symptomatic, additional diagnostic testing is necessary regardless of the NBS result. Despite the algorithm used to determine infants at highest risk, NBS may not detect all affected newborns.\(^2-5\) A positive or abnormal screening result is not a confirmed diagnosis of a disorder, and a negative or normal screening result is not a confirmed exclusion of a disorder.

2. Background

Newborn Screening (NBS) is a public health program that identifies infants at risk for certain conditions that may lead to death or disability if undetected or left untreated. NBS includes analysis of biochemical or molecular markers in dried blood spots (DBS) conducted at a laboratory in addition to point-of-care screening performed at birthing facilities (i.e., hearing and critical congenital heart disease (CCHD) screening). This document will focus on the analysis of biomarkers in DBS.

The DBS specimen is an inherently complex matrix, with variables including hematocrit and leukocyte counts that can have broad effects on screening results and cutoff values. Additionally, for screened conditions, there is a clinical spectrum of phenotypes that can impact the biomarker levels and the laboratory’s ability to detect the disorder in the newborn period. In general, marker levels of newborns with mild phenotypes are closer to the marker levels in the reference population than those with classic phenotypes. Establishing an algorithm to detect mild cases may increase the false positive rate (decreasing the positive predictive value) of the screen—an important consideration when balancing the benefits and harms of screens and the impact of an abundance of false positive results.

Newborns with positive screen results are referred for diagnostic testing. In some cases, an additional specimen is requested to repeat the screen. DBS screening, like any screen, may not identify every instance of disease as pre-analytical and analytical variables described in section 2.1 below can impact the results.
2.1 Cutoffs

Traditionally, cutoff values have been used to determine if a newborn is at risk for a disorder; cutoffs can be quantitative, semi-quantitative or qualitative. A cutoff may also be at the low or the high end of the reference range depending on the biomarker being evaluated and the disorder being screened. The cutoff may identify a reduction of analyte (i.e., biomarker is low) or an accumulation of analyte (i.e., biomarker is high). In newborn screens, cutoff values can also be fixed or floating. A fixed cutoff value is based on the analytical result. A floating cutoff is typically based on a percentile, such as the top 5% of biomarker levels/screening results obtained on a particular day or using a particular reagent lot. Other examples include the multiple of median (MoM) method and the percent of mean method, which can take into consideration the day on which the screen is performed, the instrument used and/or the population mean of screening results.

To establish cutoffs, the screening laboratory must first establish a reference range for the disease marker(s). The reference range is the range of marker concentrations that would be expected in a healthy person measured by the laboratory for the specified method. Due to analytical and other variables listed below, reference ranges and subsequently determined cutoffs will vary between NBS programs, and must be established or verified by each individual laboratory for each test performed.\(^{6,7}\) The variability alone does not affect the performance of the screen. Each screening laboratory should set cutoffs based on the expectations of the screening assay, while considering the phenotype of the disorder being screened and the sensitivity and selectivity of the screen. Programs should also determine their tolerance for generating higher false positive rates to reduce the chances of false negatives, in addition to the factors below:

A. Age and health status of the infant at time of sample collection  
B. Birth weight  
C. Environmental conditions during dried blood spot transport  
D. The collective experience of NBS programs screening for the disorder, or if it is a disorder that is not yet widely screened nationally  
E. The specific method used for the analysis  
F. Instrument platform  
G. Sensitivity of the instrument  
H. Stability of the measured analytes  
I. Availability of reference materials  
J. Calibration material or the internal standard used  
K. Target condition (time critical and severe to mild cases)  
L. Program policies with respect to the screened conditions  
M. The laboratory’s screening algorithm including:  
   1. Use of multiple markers to determine screen results  
   2. Acceptable number of false positive results  
   3. Race and ethnicities within the target population  
   4. Use of second-tier screening test(s)

Second-tier screening is necessary with some screening tests if the primary marker used for screening yields a high false positive rate, as is the case with cystic fibrosis screening using the primary biomarker
immunoreactive trypsinogen (IRT). To decrease the false positive rates, newborn screening programs may implement second-tier screening, which provides a more selective screen and may improve positive predictive values. Second-tier screens use the same dried blood spot that was collected for the first-tier screen and are valuable as no additional patient contact is necessary.

The factors listed above and others will affect the reference range, which includes the calculated mean, median and other statistically derived values of biomarker concentrations measured in a population study. Additional variables to be considered when setting newborn screening cutoffs include:

A. Age at time of specimen collection  
B. Feeding status  
C. Gestational age/ prematurity or birth weight  
D. Transfusion status  
E. Gender or race or ethnicity (for select conditions)  
F. Whether the NBS program collects one or two specimens per newborn  
G. Analyte results from initial instrument validations, including true positive and false negative cases and quality assurance results from the platform to be used  
H. Instrument variability and whether instrument-specific cutoffs versus method-specific cutoffs will be utilized  
I. Nature of the screening test (i.e., enzyme or immunoassay, vs. direct measurement of biomarker concentration such as in tandem mass spectrometry (MS/MS) for amino acid disorders)
3. Overview of Cutoff Determination

Generally, a cutoff value may be determined by: 1) testing dried blood spot (DBS) specimens from unaffected newborns and using manufactured controls (made in-house, purchased commercially, or obtained from the Newborn Screening Quality Assurance Program, Atlanta, GA); 2) analyzing data to determine whether the screen has adequate precision and accuracy; 3) assigning a preliminary cutoff value; 4) validating or verifying the cutoff using known positive control samples, preferably from residual DBS, when available, and; 5) comparing cutoff values to other programs (directly with other state NBS programs or via Region 4 Stork Collaborative Project database). The steps below summarize the process for determining cutoff values.

3.1 Testing Dried Blood Spots from Newborns

Perform a population study using manufactured controls and DBS specimens from presumed unaffected newborns. Obtain residual DBS that offer minimal limitations related to storage conditions and age (i.e., recently obtained residual DBS perform better than their older counterparts). Sample size for a population study will vary depending on several factors, including a laboratory’s methodological guidelines, confidence that the data is representative of the population and/or the method’s run-to-run variability, the state’s birth rate, disease incidence and the screening implementation timeframe. The general practice is to analyze hundreds to thousands of specimens. Larger sample sizes may be needed to evaluate birth weight, age, seasonal effects and other variables. Specimens should be tested on all instruments that will be used to perform routine screening to understand and compensate for instrument-to-instrument variability prior to the start of screening (instrument matching protocols should be in place).

3.2 Analyzing Data and Determining a Preliminary Cutoff

A. Evaluate whether the precision of the method is adequate to differentiate results close to the cutoff (see B.1. below). Consider running a higher sensitivity method or more replicates (punches from same DBS) to ensure precision of the final result. For qualitative tests, determine percentages of results identified as normal and abnormal before setting a preliminary cutoff. Test true positive controls only after the precision and accuracy of the methods have been evaluated and deemed acceptable.

B. For quantitative or semi-quantitative tests, use statistical software such as Excel, Statistical Analysis Software (SAS), Laboratory Information Management Systems (LiMS) vendor software or others to calculate the mean, median, standard deviation and possibly the percentiles of the data set. These can also be calculated using the log transformed data.
   1. Remove outliers prior to the calculations.
   2. Using the manufactured controls, evaluate the precision of the assay (and the accuracy if possible). Generally, the precision is most important with an acceptable coefficient of variation (%CV) of < +/- 20% for the simulated positive controls.
   3. If using log transformed data, the antilog of the statistic can be used to characterize the reference range and cutoff.
   4. The mean or median plus a number of standard deviations can be calculated and used to represent the reference range (e.g., the median plus 3 standard deviations).
5. View the data graphically by creating a frequency histogram or probability density function; for example, plot the x-axis with the range of analytical values and the y-axis with the number of specimens.

6. Check population data to determine if it has a normal distribution (e.g., mean/median = 1.0).

7. Determine if demographic factors impact the distribution, such as birth weight, gestational age, Neonatal Intensive Care Unit (NICU) status, transfusion status, feeding status or the infant’s age at time of specimen collection. For example, this can be done by filtering out all specimens from infants with birthweights < 1,500 g, calculating the mean, median, standard deviation of the data set and ranges and comparing to the normal birth weight population. Also compare histograms for each sub-population. Analysis of variance (ANOVA) methodology can be used to determine if there are statistically significant results between sub-populations. If the data is statistically different, consider using the variable (e.g., birth weight) in setting the cutoff(s).

8. The preliminary cutoffs can be set using percentiles (e.g., > 99.9 or other percent), number of standard deviations above or below the mean or median as compared to other laboratory results, or by choosing a fixed or floating cutoff that is expected or demonstrated based on comparison of the results to the manufactured controls or other available information (e.g., from diagnostic laboratory results, where comparison of affected and unaffected ranges are already established).

9. A preliminary conservative cutoff can be set by comparing the normal and positive sample results (manufactured positive controls or other positives that were tested in the population study). A conservative cutoff should allow for detection of all true positives, and should not create a projected screen positive rate that is high relative to the expected incidence rate for the associated condition.

3.3 Validating or Verifying the Cutoff

A. Challenge the preliminary cutoff
   1. When possible, test known true positive, false positive, carrier, pseudodeficiency and poor quality/diluted specimens from stored residual NBS specimens. True positive samples or blinded patient panels may also be requested from other laboratories.
   2. For some disorders, true positive specimens from older patients can be used (e.g., GALT or Cystic Fibrosis mutation), but only with a thorough understanding of the disorder and how the marker is affected by patient age. Specimens from older patients may not yield similar results to the values determined in newborns.
   3. Test proficiency specimens from the CDC Newborn Screening Quality Assurance Program (NSQAP)\(^8\) and compare the results to those obtained by other NBS programs.

B. Compare the cutoff
   1. Contact other laboratories that are using a similar method.
   2. For laboratories using an FDA-approved kit, compare the calculated cutoff to the cutoff that is provided in the package insert.
   3. The Region 4 Stork (R4S) Laboratory Performance Database and Clinical Laboratory Integrated Reports (CLIR)\(^9,13\) can be used to compare population reference data and cutoffs from other programs for tandem mass spectrometry (MS/MS), lysosomal storage disorders (LSDs) and adrenoleukodystrophy screening methods. Note that R4S is
currently available, but this software will be phased out in the next couple of years and replaced by the second generation of the software, CLIR.

C. Conduct a literature search to identify published cutoffs or reference ranges and the prevalence and incidence of the disorder. This information can be used to estimate the screen positive rate and be compared to the incidence rate.

D. Evaluate results of the population study
   1. Compare results from true positive specimens to results from the population study.
   2. Determine the proportion of true positive specimens tested as part of the study that would be detected at the chosen cutoff. Ideally, all positive specimens will be detected. If the results are not acceptable, modify the cutoff so that all will be detected; or consider the quality of the true positive specimens, certainty of diagnosis and associated phenotype prior to adjusting the cutoff. True positive specimens with results that are close to the cutoff can be most informative and are useful to determine if the precision of the method is acceptable at the corresponding samples marker level.

E. Set the cutoff
   1. Determine whether the cutoff will be floating or fixed (see below section “Special Considerations”). If other laboratories are using the same assay, consider if they are using fixed or floating cutoffs.
   2. Consider the projected number of screen positive newborns at risk for the disorder and consider consulting with your medical providers to ensure this number is acceptable to them.
   3. Consider setting conservative cutoffs initially to minimize the possibility of false negative test results. However, also consider that if a conservative cutoff is chosen, the screen may lead to the detection of newborns with mild or late onset disease that may or may not develop until later in life, such as with LSDs. Consider if this is an acceptable component of reported results for your program.
   4. Working with clinical specialists, closely monitor the follow-up results of screen positive infants. Is there any correlation of clinical results with screen results (degree of out-of-normal range)? Is phenotype mild or severe? Are there other factors that may have affected the screen results such as prematurity or age of the newborn at the time of sample collection? Continue to collect as much follow-up data on the referred infants and work with the clinicians and advisory board to adjust the cutoffs if necessary.

Cutoffs are not static; they should be monitored and adjusted based on feedback from short- and long-term follow up and clinical outcomes of infants, when new information about the disorder becomes available or when significant changes are made in the testing process that may affect the reference range. This includes events like revisions to analytical methods, instrument replacements and/or new substitutions to key consumables or chemicals used during testing.

3.4 Special Considerations

A. If the laboratory is the first to develop a method to screen for a new disorder, setting cutoffs is particularly challenging.
   1. The quality of the positive patient specimens is important. Use original newborn DBS specimens, if available.
2. Consider the phenotype of the individuals used as positive patient control specimens. Usually, positive specimens from individuals with a classic or severe phenotype of the disease are available for testing. Specimens from patients with milder phenotypes may not have similar marker values. Prior to establishing a cutoff value, gather input from specialists to determine if mild or late-onset phenotypes should be identified.

3. Perform a literature review to determine diagnostic methods currently used, normal and abnormal reference ranges and the sample matrix used for diagnosis. Often, the same biomarkers are used in diagnostic testing and newborn screening (NBS), and the associated reference ranges used in the diagnostic test can be a useful guide in setting the NBS cutoffs.

4. Determine where the marker resides in the blood (serum, red cells, or white cells). This will help predict the potential for variability on the measured values and differences in diagnostic versus screening reference ranges. For example, when testing for galactosemia using an assay that measures GALK enzyme activity, GALK is present only in red cells, so a negative screen may be a false negative if the newborn was transfused with red cells prior to obtaining the DBS specimen.

B. If the test has already been well-established by another laboratory, consider performing an exchange with a panel of normal, abnormal, and, if applicable, carrier or pseudodeficiency specimens. Reference ranges and results of the specimen exchange can be used to determine concordance between the two data sets. For example, if there is systematic bias observed in the comparison test, the results can be used to extrapolate reference ranges for the varied cohorts and in setting cutoffs, even if very few positive or other controls were used in the comparison.

C. Consider using borderline cutoffs. Often, if the condition is not imminently life threatening, NBS programs will set borderline cutoffs. A borderline cutoff is a range of values between normal and referral level marker concentrations. If a specimen is found to be in the borderline range, another specimen is requested rather than referring the newborn for follow-up diagnostic testing. The borderline cutoffs are helpful when the marker being measured may increase or decrease depending on the age of the newborn at the time of specimen collection, or if the marker tends to be elevated at birth due to stress on the newborn resulting from the birthing process (such as 17-OHP). The CDC is in early stages of developing quality assurance dried blood spot (DBS) materials that mimic historical confirmed cases that were close to NBS cutoff values. The intent of this initiative is to provide NBS programs with positive controls that produce results in the borderline range. Often, there are few true positive specimens that can be shared with other laboratories to challenge the cutoff, and there is limited residual specimen for those that do exist. These CDC produced DBS will allow laboratories to challenge their cutoffs.

D. Determine whether to use fixed or floating cutoffs. Generally, fixed cutoffs are used for assays that directly measure marker concentrations (e.g., MS/MS analysis of phenylalanine). Whereas floating cutoffs are used for indirect or functional assays, such as enzyme, or immunoassays where the concentration of the marker in question is not directly measured by the instrumentation and is dependent on an enzyme reaction or binding of an antibody to an antigen. With the latter assays, more uncontrolled variables can affect the day-to-day or lot-to-lot performance of the assay and the use of a floating cutoff compensates for these variables. With floating cutoffs, it is important to consider if variability of the assay is more day-to-day or lot-to-lot, and if the daily or population mean/median will be used. For example, using percent
of daily mean or median would be more valuable if there is daily variability. The use of median may be preferred over mean due to it being less affected by outliers in the data set.

Some laboratories use Multiples of Median (MoM), which can be thought of as application of a floating cutoff since a percentage is referenced. However, the general intent for a floating cutoff and using MoMs is nearly the opposite. A floating cutoff is typically used where large assay variation is known; using a percentage provides a comparable reference range from day to day. MoMs are typically used where assay variation is minimal so that reference to the median shows when assay performance has changed.

E. Based on the information in this guideline and other resources, it is recommended that NBS programs have a general standard operating procedure that reflects the local requirements for establishing and for performing ongoing evaluation of method cutoffs.

3.5 Collaborative Laboratory Integrated Reports

Collaborative programs such as Collaborative Laboratory Integrated Reports (CLIR) can also be used to provide additional risk assessment for a program’s chosen cutoff values. CLIR is a novel approach developed by the Biochemical Genetics Laboratory at the Mayo Clinic that uses a scoring system based on condition-specific degree of overlap between covariate-adjusted reference and disease ranges. Scores are integrated for all informative markers and calculated ratios by post-analytical interpretive tools based on multivariate pattern recognition software, and provides an alternative to absolute cutoff values with a single threshold of likelihood of disease. Adjustments are calculated by regression models able to factor in two continuous covariates (typically, age at collection and birth weight), one categorical covariate (sex) and also harmonization of data by location. The computational approach is to compare each marker to its reference range adjusted for the birth weight and age at collection of the patient, and then to calculate a score based on the degree that each result penetrates the disease range above or below the corresponding reference interval. The software relies on large amounts of data (millions of data points are ideal). Current data sets are adequate from one hour up to one year of age and a birth weight >500 g.

Newborn screening programs that enter data may use CLIR in a variety of ways:

A. CLIR can be used to prospectively analyze screening data. High throughput functionalities (tool runners) are available in CLIR to upload data and return a report identifying patient specimens at variable levels of risk (possible, likely, very likely) for each of the screened conditions, allowing programs to focus their attention on fewer cases, allowing greater efficiency.

B. CLIR can be used to identify analytes and ratios that are informative for a disorder. CLIR includes tools that identify analytes and ratios where the separation between reference and disease range is clinically significant, meaning a degree of overlap less than 50%. This tool allows the laboratory to review analyte ratios that separate normal and affected specimens.

C. CLIR can assist in reducing false positive results. Tools are also available to provide a differential diagnosis between target conditions and particularly between true and potential false positive cases (for example, carriers manifesting a milder biochemical phenotype).

D. CLIR allows for laboratory comparisons. The interpretative tools for each condition plot the values of scores for the cases in three groups: the user’s site, a choice of aggregator (e.g. the United States) and the entire project worldwide. This allows the laboratory to compare a
particular specimen to other cases, and allows visual comparison of the cases from one laboratory to others.

E. **CLIR can provide indications for differential diagnosis.** CLIR tools allow laboratories and follow up centers to compare the extent to which a particular profile matches any of a number of conditions characterized by particular elevations.

CLIR offers users the following functionalities:

A. **Covariate adjustment.** Analytes and ratios are adjusted for demographic variables such as birth weight and age at sample collection. In addition, sex differences are assessed and, if needed, a further adjustment is applied.

B. **Location harmonization.** Allows for direct comparison of values or cutoffs across contributing laboratories.

C. **Global contribution of diagnosed case data.** Results in a compilation of positive cases to help determine the distribution of markers in the affected population.

D. **Large database of normal profiles.** The database of normal profiles allows the incorporation of new tools and/or new analyte combinations so the normal distribution can be evaluated on existing data.
4. Disorder Specific Considerations

4.1 Endocrine Disorders (Congenital Hypothyroidism, Congenital Adrenal Hyperplasia)

Screening for endocrine disorders is challenging as there are significant dependencies affecting the marker concentrations used in screening for these disorders including birthweight, gestational age and age at time of sample collection. Consider these variables when setting cutoffs.

A. Box plots may be useful if the cutoffs need to be tiered by birth weight or age at time of specimen collection (e.g., box plots created for weight ranges can show birth weights where values change significantly).

B. 17OHP cutoffs are typically tiered by gestational age or birth weight.\(^{(16)}\)

C. TSH cutoffs are typically tiered by the age of the infant at the time of specimen collection. This in addition to a Low Birth Weight Protocol helps to identify delayed cases of congenital hypothyroidism in premature infants.\(^{(20)}\)

Since endocrine disorders have been on state NBS panels for a long time, it is beneficial to consult the literature and work with other programs to determine how they are setting cutoffs for these disorders.

4.2 Cystic Fibrosis

A. Screening for cystic fibrosis is primarily performed with immunoreactive trypsinogen (IRT) and second tier DNA analysis for mutations in the CFTR gene. DNA analysis is performed on specimens with an elevated IRT. NBS programs use these analyses in combination with testing algorithms such as IRT/IRT, IRT/IRT/DNA, or IRT/DNA. IRT is the marker used in the primary screen and because of the high false positive rate with this marker, second-tier testing is recommended.

B. The IRT cutoff is floating and/or fixed. A floating cutoff is recommended because IRT is subject to seasonal variations and lot-to-lot variability of the reagents.

C. A floating cutoff is calculated daily. Laboratories have set this cutoff at a percentage between the 95th and 99th percentiles to identify the daily batch of specimens with high IRT values that require second-tier testing. Use at least 100 specimens for an appropriate percentile calculation with a floating cutoff.

D. A fixed cutoff can be set that corresponds to the 95-99th percentile if a floating cutoff is not feasible. A fixed cutoff can also be added as a failsafe, if the floating cutoff is higher than expected in that particular batch of specimens.\(^{(10)}\)

E. A borderline cutoff can be set for ultra-high IRT values where no mutations are identified during DNA analysis. A borderline result initiates a repeat specimen collection for the infant. Two elevated IRT results from two different specimens on the same infant indicate further diagnostic testing is required.

F. Neonatal intestinal obstruction due to meconium ileus is one of the signs of cystic fibrosis that occurs in approximately 20 percent of affected newborns. This can result in an IRT value within the normal range. In these instances a sweat chloride test, typically reserved for diagnosis following a positive screen, is recommended despite an IRT value within the normal range.\(^{(10)}\)

G. Blood level IRT concentration is fairly stable until about 2 weeks of age when it begins to decrease. Age related cutoffs are recommended for infants greater than 7 days of age.

H. IRT concentrations in premature, low birth weight, or sick infants may be high due to complications other than cystic fibrosis. Additionally, African American infants have higher mean...
IRT concentrations compared to White and non-Hispanic infants. This often results in more false positives in these populations. (10)

4.3 Hemoglobinopathies

Most hemoglobin disorders detected through newborn screening are detected qualitatively and do not require cutoffs as in the case of sickle cell and sickle C diseases. Alpha thalassemia, however, is detected through Hemoglobin (Hb) Bart’s quantitation as a percentage which is directly proportional to disease severity. Bart’s, along with acetylated and glycated forms, is a gamma chain tetramer formed when there is a reduction in functional alpha chains. The best way to establish cutoffs to determine alpha thalassemia severity based on Bart’s percentage is to perform a phenotype/genotype correlation. When that is not possible, cutoffs may be established by comparing Bart’s percentages through a sample exchange with another program with expertise in reporting alpha thalassemia.

Isoelectric focusing, high performance liquid chromatography (HPLC), or a combination of both may be used to detect Hb Bart’s as described below.

A. Isoelectric focusing (IEF) concentrates hemoglobin proteins based on their isoelectric charge through electrophoresis using an agarose gel medium infused with carrier ampholytes. Since Hb Bart’s is detected visually the quantitative interpretation is subjective, and differentiating the severity level of the alpha thalassemia (silent carrier, alpha thalassemia trait, and Hb H disease) is challenging using IEF alone. Some labs use a densitometer to quantify the IEF percentage which improves accuracy.

B. HPLC is the preferred method for hemoglobin percent quantitation due to its precision and accuracy. It is important to note that different HPLC systems may have different Bart’s percentages depending on resolution. Also, percentage variation can occur between column resin lots. The majority of NBS programs report Hb H disease (three affected genes) which corresponds to 15 – 30 percent Hb Bart’s depending on HPLC resolution.

4.4 Enzyme Deficiency Disorders

Screening tests for galactosemia, biotinidase deficiency and lysosomal storage disorders (LSDs) can use fixed or floating cutoffs. Often floating cutoffs are used with these functional assays due to seasonal and reagent variations that affect the calculated activities. The following have been used for floating cutoffs:

A. A multiple of the median (MoM) based cutoff can be calculated from the data by comparing results of positive screens to the median of the population rather than a daily calculated median.

B. Percent of the daily mean or percent of daily median based cutoffs can be calculated from the data. The percent of mean or median cutoff is set to detect all known positives; generally, it is set conservatively at a value considerably higher than the highest true positive. Note that the daily median is less prone to the effects from high activity outliers. Often in enzyme testing, the measured activities of a few samples will be highly elevated and skew the mean; use of the median can minimize this effect.

If using a floating cutoff, be aware that the results from low birthweight, early collection and sick infants may skew the mean or median values for the day. Excluding these sample types in calculation of the
mean or median used in the floating cutoffs will lead to stabilization of the results. This is especially true for states with low sample volumes.

Since lysosomal storage and enzyme deficiency disorders are screened by measuring the product concentration near the low end of the reference range, it is important to consider the precision of the test near the cutoff. This can be accomplished by taking multiple measurements (more than in duplicate) and averaging the results for comparison to the cutoff.

Screening for lysosomal storage disorders (LSDs) is highly dependent on the variables of the newborns’ birthweight and age of the newborn at the time of sample collection. Sample storage and transport conditions can also affect measured enzyme activities. These variables should be considered in evaluation of cutoffs. The CLIR tool can be especially helpful in evaluating the risk of the newborn for a screened LSD, as the tool relies on the covariates of age and birthweight in establishing normal ranges for the LSDs.

4.5 Amino Acid, Fatty Acid Oxidation and Organic Acid disorders

A. Amino acid and acylcarnitines are measured using tandem mass spectrometry (MS/MS). CLSI NBS04-02, Newborn Screening by Tandem Mass Spectrometry establishes best practices for analyzing markers and setting cutoffs using this method; this document is recommended as a reference.

B. Consult specific method and instrument parameters to optimize performance.

C. If a metabolic profile will be used to interpret results rather than a single marker, assess cutoffs for all markers, including any ratios used.

D. According to the CLSI NBS04-02 recommendations, if a program analyzes specimens collected outside of the newborn period (greater than 7 days of life), validate separate age-specific cutoffs.

E. When calculating population study data statistics, consider developing two data sets to analyze and compare the statistical parameters. Include data points from all patients for one data set and exclude data from premature, low birth weight or neonatal intensive care unit (NICU) infants for the other data set. The use of total parenteral nutrition (TPN) and medications in this special population can elevate some amino acid and acylcarnitine levels. If a significant difference is noted in the statistical parameters, setting separate cutoffs may help reduce the risk of false negatives in the normal population and false positives in the special population.

F. Fixed cutoffs may be assigned using means and standard deviations, or percentile data.

4.6 Severe Combined Immunodeficiency (SCID)

SCID is screened for by quantitating T-cell receptor excision circles (TRECs) in extracted DNA from dried blood spots.

A. SCID can use a floating cutoff such as the MoM approach to assess the amount of TREC and reference gene in the dried blood spot punch.

B. A fixed cutoff can be determined using calibration standards to quantify the number of TREC and reference gene copies per microliter (µL) of blood.

C. A fixed cutoff can also be based on a cycle of quantitation (Cq) value rather than a TREC copy number.
D. Screening for SCID is challenging in sick and premature infants as they can have low levels of TREC and not have SCID. This screening also identifies other t-cell lymphopenias.\(^{(20)}\)

4.7 X-Linked Adrenoleukodystrophy (X-ALD)

X-ALD is screened for by quantitating lysophosphatidylcholines (LPCs) in dried blood spots using either high profile liquid chromatography (HPLC) tandem mass spectrometry (MS/MS) or flow injection analysis (FIA)-MS/MS.

A. This can be a fixed cutoff for multiple LPCs such as C20:0-LPC, C22:0-LPC, C24:0-LPC, and C26:0-LPC.

B. When using flow injection positive ion MS/MS analysis to screen for X-ALD, there is an unknown isobaric interference in determining the concentration of C26:0-LPC. Since the deuterated internal standard does not correct for effects of this isobaric interference, instrument-to-instrument and day-to-day concentrations of C26:0-LPC will vary more than is observed with other MS/MS assays (AA/AC). Consider adjusting the daily median or mean for all results to match the population mean, so that the individual sample results and the daily mean for this marker are consistent with the results measured during the validation study. The relative response factor can be adjusted as needed to achieve consistency in daily or instrument-to-instrument variability. Another option would be to consider a floating cutoff for this marker, but this would need to be done for each instrument, as instrument performance can vary day to day, and is common after performing routine maintenance.

C. Additionally, fixed cutoffs can be established for ratios between the LPC markers, for example, C24:0 LPC over C20:0 LPC or C26:0 LPC over C22:0 LPC as well as other LPC ratios may be useful.
5. Monitoring and Ongoing Evaluation of Cutoffs

A. Monitor cutoffs closely for the first year by evaluating the test sensitivity and specificity when possible. Use the number of borderline and positive results, false positive rate and positive predictive value in the evaluation based on the follow up outcomes.

B. Determine if the cutoffs are identifying the expected number of positive cases for your state’s population and if the false positive rate is reasonable or high based on the expected and detected incidence. Typically, once conducting routine screening, these adjustments will be made based on feedback from the follow-up program (e.g., if primarily mild phenotypes are detected at the present cutoff, etc.).

C. Meet with specialists, if possible, to evaluate population data. Factors to evaluate include review of new true positive cases, false positive cases, false negative cases and any new peer-reviewed literature. Programs may also consider comparing their cutoff values and risk assessments to CLIR or other newborn screening programs.

D. Consider re-evaluating cutoffs following kit changes, changes in equipment or test modification or when new clinical information is received (e.g., from cases, knowledge of natural history of the disease).

E. Check the cutoffs every six months when routine screening begins. (21)

F. In the event that the screening laboratory has been informed of a false negative result, it is important to fully investigate the finding. To make this determination:
   1. Obtain the diagnostic results and consult with the specialist to understand the exact diagnosis, if the infant is truly affected by the disorder being screened for and if it is a disorder that should be detected by the screen.
   2. Repeat the screen to see if the original result is obtained. If the second screen result is within experimental error of the original result (as defined by the NBS program), and the result is far removed from the current cutoff, check to see if any other extenuating circumstances may have affected the measured result such as sample quality, birthweight, delays in sampling, transport or feeding issues. Another possibility may be that the sample is from a different infant.
   3. If all other factors have been ruled out, consider adjusting the cutoff or testing the algorithm to reduce screen negative results while not introducing an unacceptable number of false positive results.
   4. Consider if there are other markers or ratios of markers that can be used to detect the sample as screen positive.
   5. If available, consider using CLIR to evaluate the sample to determine if it would have been flagged as at risk for the condition.
   6. Send the sample to another screening laboratory to ensure consistency of results and for cutoff comparisons. Determine if other laboratories would have detected this sample as a screen positive.
6. References


(9) Laboratory Performance Database (R4S) [Internet]. [cited 2017 Dec 11]; Available from: https://www.nbstrn.org/research-tools/lab-performance-database


(13) Collaborative Laboratory Integrated Reports (CLIR) [Internet]. [cited 2017 Dec 11]; Available from: https://clir.mayo.edu/Home/About


§ 493.1281 Standard: Comparison of test results