FDA Executive Summary

Prepared for the
August 10, 2016 meeting of the
Clinical Chemistry and Clinical Toxicology Devices Panel
DEN150035
Seeker System
Baebies, Inc.

Introduction

This document is the **FDA Executive Summary** for the meeting of the Clinical Chemistry and Clinical Toxicology Devices Advisory Panel meeting on the Seeker system from Baebies, Inc. The sponsor (Baebies) has submitted a de novo application (DEN150035) to market their newborn screening test system. The Seeker system is intended for quantitative measurement of the activity of multiple lysosomal enzymes from newborn dried blood spot specimens. Reduced activity of these enzymes may be indicative of a lysosomal storage disorder. The enzymes measured using the Seeker™ Reagent Kit and their associated lysosomal storage disorder are listed in the following table:

Enzyme (abbreviation)	Disorder
α-L-iduronidase (IDUA)	Mucopolysaccharidosis Type I (MPS I)
α-D-glucosidase (GAA)	Pompe
β-glucocerebrosidase (GBA)	Gaucher
α-D-galactosidase A (GLA)	Fabry

Reduced activity for any of the four enzymes (indicative of a lysosomal storage disorder) must be verified by other confirmatory diagnostic methods. The submission (DEN150035) is under review by the Division of Chemistry and Toxicology Devices (DCTD), Office of *In vitro* Diagnostics and Radiological Health (OIR), within the Center for Devices and Radiological Health (CDRH) of the Food and Drug Administration (FDA).

This document will provide background information on newborn screening and the four lysosomal storage disorders that this device will screen for, describe the results from a statewide pivotal validation study, and the analytical performance validation of the system, and summarize the areas for which FDA seeks expertise and input from the Clinical Chemistry and Clinical Toxicology Devices Advisory Panel. FDA seeks input on whether this device should be authorized for marketing, and if so, how to accurately communicate test performance to laboratories who may want to implement this testing. FDA also seeks the Advisory Panel's opinion on whether age specific cutoffs should be incorporated in the statistical analysis of the study and/or the Instructions for Use for the test. FDA is requesting input on recommendations for addressing the potential impact of high temperature and humidity on enzyme stability of stored dried blood samples. Finally, FDA is asking for the panel's perspective on the analytical performance of the test.

Table of Contents

I. Background

Newborn Screening

Newborn screening was first introduced in the United States in the early 1960's to screen for phenylketonuria and has subsequently expanded to include many more conditions. Dried blood spot based newborn screening is typically conducted by state health departments, and each state determines what conditions will be screened. Newborn screening programs are designed to identify conditions in newborns that are treatable, but not clinically evident at birth, for which intervention prior to diagnosis may improve clinical outcomes. The U.S. Department of Human and Health Services' Advisory Committee on Heritable Disorders in Newborns and Children (ACHDNC) was established by the Newborn Screening Saves Lives Act of 2007 and its mission is to reduce morbidity and mortality in newborns and children who are at risk for heritable disorders. The ACHDNC recommends that every newborn screening program include a recommended uniform screening panel (RUSP) that currently lists 32 core disorders and 26 secondary disorders, and the HHS Secretary has adopted the recommendations of the ACHDNC.

Almost all newborn screening tests use dried blood spots (DBS) as the specimen. Healthcare professionals collect dried blood spot specimens by applying a few drops of freshly drawn blood from a heel stick onto filter paper. The specimens are air dried and shipped to public health laboratories where they are analyzed using different methods depending on the conditions being screened by that state.

A screening test is used to evaluate asymptomatic patients to identify patients who may have (or be at risk for developing) a particular disease or condition and who should receive intensive follow-up testing for confirmation. The newborns identified by newborn screening are then sent for diagnostic testing to definitively determine if they have (or are at high risk for having) the disease (confirmed positive), or whether they are unaffected for that condition (false positive). The goal of a screening program is to have a very low false negative rate given that there may be no other opportunity to identify these newborns before they have developed a serious disease; many of the conditions screened for could have significant morbidity or even mortality. The false positive rate should also be low; programs would like to limit the number of newborns sent on for confirmatory diagnostic testing, both because of the anxiety the families experience as well as the programs' resources. Diagnostic testing typically uses tests with high sensitivity and specificity; often a panel of tests may be used for diagnostic testing, along with an assessment of clinical signs and symptoms, to determine if a newborn is confirmed positive.

Diseases Screened Using the Seeker System

Lysosomal Storage Diseases (LSDs) are a large group of disorders caused by a deficiency of a specific enzyme responsible for the degradation of substances in lysosomes. Lysosomes are contained in all cells (except for red blood cells) and thus this set of diseases can affect different organs and systems at the same time. The clinical picture includes various forms of mental retardation (i.e., MPS 1,Gaucher), liver/spleen enlargement, bone abnormalities, kidney failure (i.e., Fabry), cardiac diseases, corneal clouding (i.e., Fabry and MPS I), and muscle weakness (i.e., Pompe)[1-5].

Treatments for these conditions include enzyme replacement therapies (ERT) and hematopoietic stem cell transplant.

Depending on the severity of the disease, the chronicity of the diseases vary. Infantile forms of Pompe are devastating with poor survival. The classic infantile form (with Cardiomyopathy) usually leads to death within

the first year of life without treatment[4,6]. The Non-classic infantile form (without Cardiomyopathy) has longer survival, but without treatment, death usually occurs in early childhood[4]. Survival is longer in juvenile or adult forms of the disease. Early diagnosis and prompt Enzyme Replacement Therapy (ERT) can affect survival. In ACHDNC's letter to the HHS Secretary, the committee wrote that Pompe screening results in earlier diagnosis and treatment of the infantile form of the disease. ACHDNC explained that ERT has been shown to significantly modify the course of the infantile form of Pompe disease and earlier treatment with ERT results in better outcomes for affected infants. Approximately 28% of Pompe disease cases are infantile-onset of which about 85% are Classic infantile onset[7]. About 75% of cases of classic infantile-onset are cross-reacting immunologic material (CRIM) positive[7]. CRIM negative patients can develop high titers of antibodies that neutralize ERT, leading to worse treatment efficacy.

Individuals affected with Fabry have a median survival of about 50 years[8]. Progressive renal disease ultimately develops among almost all patients with classic Fabry[9]. ERT is recommended in classically affected males with low or undetectable enzyme to slow kidney injury progression if he has not already progressed to significant renal failure[8].

Individuals affected with Gaucher have variable life expectancy. The disease can present incidentally in asymptomatic older adults or present as severe disease in early childhood. The Neuronopathic forms (Types 2 and 3) are associated with the worst prognosis. Patients can have rapidly progressive neurologic deterioration and death usually occurs before the child reaches 2 years old particularly with type 2[3]. ERT can be given to patients with Types 1 and 3[6]. ERT treatment is not expected to prevent the fatal neurologic outcome in type 2 disease[3].

MPS I patients also have variable course of illness depending on the form of disease. The Hurler form of the disease is most severe and associated with progressive neurologic decline[6]. According to the ACHDNC's letter to the HHS Secretary, the severe form of MPS I is associated with early onset of developmental delay followed by developmental regression, cardio-pulmonary failure, and obstructive airway disease, with death occurring before the age of 10 years. ACHDNC stated that newborn screening would have a significant benefit in terms of cognitive outcomes, however, the benefits of early detection on overall survival were not known due to the small sample size studies and duration of treatment within the studies.

Screening for LSDs

Interest in newborn screening for LSDs has expanded recently due to the availability of treatments and technologies to screen for these conditions. Currently, LSDs are either detected through enzymatic assays or molecular analysis. However, there are currently no FDA cleared or approved methods for screening newborns for LSDs. Meanwhile several states have mandated screening for lysosomal storage conditions within the last few years. Illinois, Missouri, New Jersey, New York, New Mexico, Pennsylvania and Kentucky have added different LSDs and combinations of LSDs to the list of conditions that they screen for. Arizona is evaluating whether it should include screening for Krabbe. Pompe and MPS I have been recently added to the RUSP by the Secretary of Health and Human Services. Newborn screening for LSDs has been performed by enzyme activity measurements in dried blood spot extracts by tandem mass spectrometry (MS/MS). Individual public health labs have developed these tests themselves using synthetic enzyme substrates since there are currently no commercially available tests for this use. Reduced activity of a particular enzyme is indicative of a risk for the corresponding condition.

II. Device Description

The Baebies Seeker System uses fluorimetry on a digital microfluidic platform to measure enzymatic activity. The digital microfluidic platform performs enzymatic analysis for newborn screening by automating all liquid-handling steps involved in an assay using sub-microliter droplets as reaction vessels. The enzymes measured and reported by the Baebies Seeker System and their corresponding lysosomal storage disorder and incidence are listed below:

Enzyme (abbreviation)	*Disorder*	*Published incidence*
α-L-iduronidase (IDUA)	MPS I	1:54,000 – 1:185,000
α-D-glucosidase (GAA)	Pompe	1:28,000
β-glucocerebrosidase (GBA)	Gaucher	1:57,000
α-D-galactosidase A (GLA)	Fabry	1:1,500 – 1:13,000

Dried blood spot extract is the specimen used to measure the enzymatic activity of the lysosomal enzymes. To prepare the dried blood spot extract, a 3.2 mm punch (containing approximately 3.1 μL blood) is extracted from the dried blood spot collected from a newborn. The dried blood spot extract is incubated in 100 μl extraction buffer for 30 minutes at room temperature. For each enzymatic reaction, 3.5 μL of extract is used.

The device and the principle of the test are described in detail in section 4 of the sponsor's Executive Summary (pages 17 to 24).

III. Regulatory History

While there are numerous devices cleared for screening newborns for several conditions (such as biotinidase deficiency, congenital hypothyroidism, congenital adrenal hyperplasia, cystic fibrosis, etc.), there are currently no FDA-cleared or approved devices indicated for screening babies for these lysosomal storage disorders.

IV. Clinical Study

A. Background

A clinical study was conducted using newborn dried blood spot (DBS) cards received at the Missouri State Public Health Laboratory (MSPHL), between January 11, 2013 and January 14, 2015. All DBS cards received at the state public health laboratory for routine newborn screening in the study period were measured using the Seeker System for the 4 LSD enzymes during the study period. No newborns were omitted from this study unless their parents opted-out of routine newborn screening on religious grounds, consistent with that state's laws. The clinical study was an investigational prospective study.

Timing of DBS Collection

DBS cards collected between 24-48 hours of life were considered optimum for newborn screening by the laboratory. If the initial DBS card was collected prior to 24 hours of life then a second (repeat) DBS card was required within 14 days of life. Initial DBS cards from ill or premature newborns were collected prior to blood

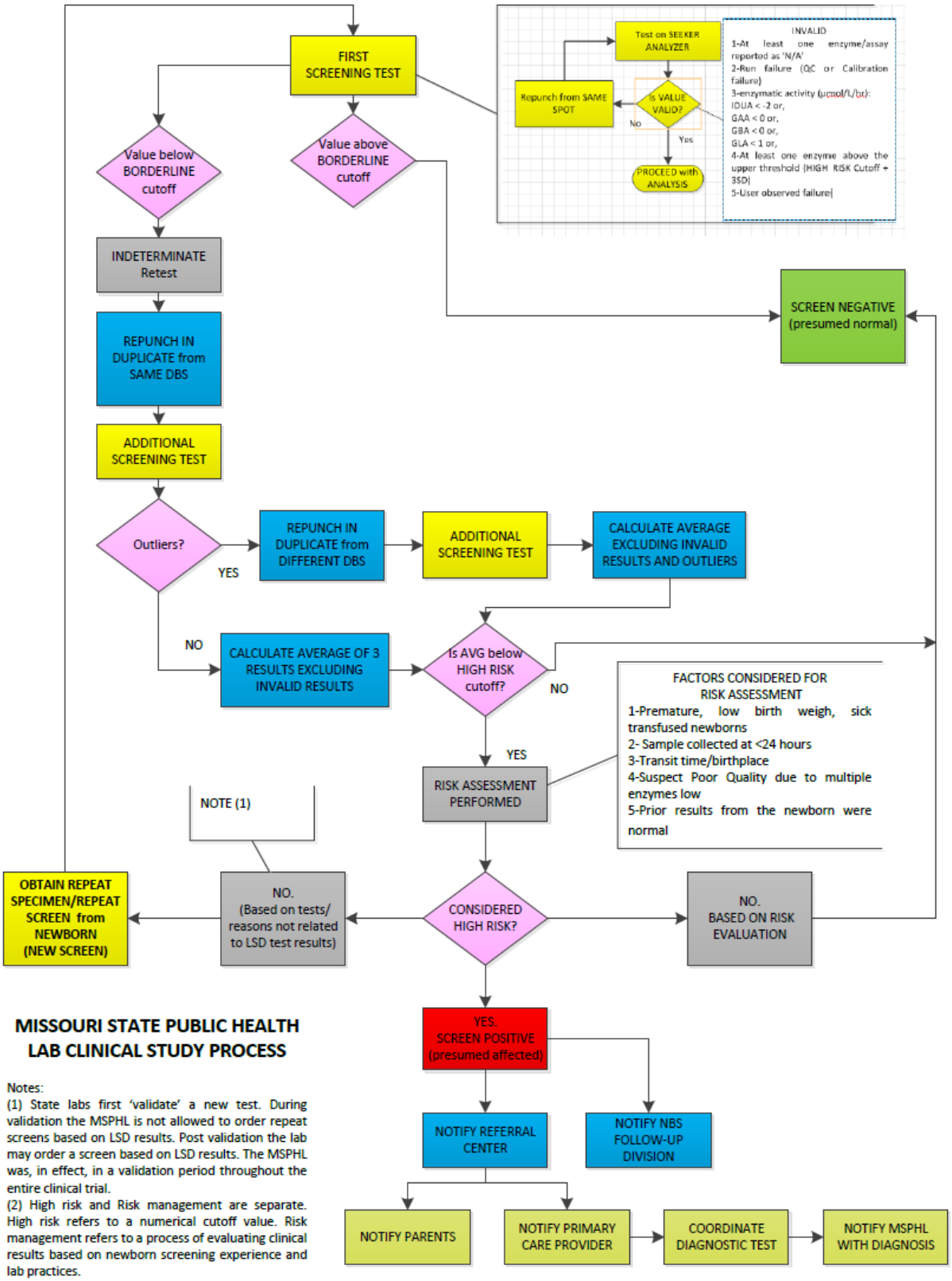
transfusion or between 24-48 hours of life. All ill or premature newborns had a repeat DBS card collected between 7-14 days of life. Lastly, DBS cards were collected at 28 days of life for infants who were less than 34 weeks gestational age, or less than 2,000 grams. Almost 95% of DBS cards were sent within 24 hours of collection to the public health laboratory via courier and delivered the next day by 7:30A.M., allowing for expedited processing.

Acceptability of DBS

All DBS received at MSPHL were examined for specimen acceptability. Poor quality specimens may not have enough blood to perform all the testing, may have been collected improperly, and or may have been delayed in the mail. Poor quality specimens were typically not tested by MSPHL (and repeat DBS cards, i.e., repeat screens, were obtained for babies with poor quality specimens). Any DBS card identified as a poor quality specimen was not included in the analysis by the sponsor.

The Screening Decision Tree

As described above, most screening programs confirm all positive results, MSPHL did not follow that paradigm for LSD screening in this study. MSPHL implemented a much more complicated scheme for determining which positive test results should be referred for confirmation testing. The screening decision tree they used is summarized below and briefly explained in the following pages:



MISSOURI STATE PUBLIC HEALTH LAB CLINICAL STUDY PROCESS

Notes:
(1) State labs first 'validate' a new test. During validation the MSPHL is not allowed to order repeat screens based on LSD results. Post validation the lab may order a screen based on LSD results. The MSPHL was, in effect, in a validation period throughout the entire clinical trial.
(2) High risk and Risk management are separate. High risk refers to a numerical cutoff value. Risk management refers to a process of evaluating clinical results based on newborn screening experience and lab practices.

- One 3.2 mm diameter punch was obtained from one blood spot from a DBS from each baby and was tested in singlicate.

- Valid run and valid test results analysis: The first step was to ensure that the run passed the quality control criteria set by MSPHL so that the run could be accepted for result evaluation. The next step was to ensure that the individual test results met the sample acceptance for enzymatic activity. Invalid samples fell into the following categories:

    - Data point was reported as "n/a"
    - Negative enzymatic values for GAA and GBA
    - GLA enzymatic activity of < 1 µmol/L/hr
    - IDUA enzymatic activity < -2 µmol/L/hr (applied only for the first test from the DBS). For example, if the first test result for IDUA was < -2µmol/hr/L, it was considered invalid. However, if upon retesting the test result consistently reported a value that was less than -2 µmol/hr/L, then the baby was considered presumed affected and referred for confirmatory diagnosis. For example, one of the newborns had an initial IDUA activity of -3.07 µmol/hr/L, which did not meet the sample acceptance cut-off for IDUA. Upon retesting the IDUA activity was -3.96 µmol/hr/L and -4.13 µmol/hr/L. This baby was referred and diagnosed as a MPS I carrier.
    - At least one enzyme (of the 4 tested by the Seeker System) that was above the upper threshold. The upper threshold was set by the MSPHL at 3 SD above the normal median for each enzyme.

  An invalid data point triggered a single re-punch and retest from the same dried blood spot.

Screening Algorithm (see Figure 1 below)

For these disorders, affected babies have low enzyme activity. In order to screen the babies, 2 cutoffs were established; a high risk cutoff and a borderline cutoff. Babies with enzymatic activity below the high risk cutoff have a high likelihood of having the disorder while babies with enzymatic activity below the borderline cutoff were considered to have an undetermined risk. Babies with enzymatic activity above the borderline cutoff were determined to be at low risk for the screened condition and presumed normal.

1. Each baby's first, valid test result was compared to the borderline cutoffs for each condition. All valid test results above the borderline cutoffs (and below the upper threshold) were considered presumed normal.

2. Any DBS with at least one enzyme below the borderline cutoffs were retested in duplicate the following work day, with a new punch from the same blood spot as the initial test. The average of the 3 results (initial results and retest in duplicate) was calculated.

    a. If the average of the 3 tests was above the high risk cutoff, the data was visually assessed for obvious outliers. If there were obvious outliers the test was repeated on an additional 2 punches from a different DBS from the same card. Note that there are 5 DBS in one dried blood spot card - all collected from the same newborn at the same time. The repeat testing algorithm is repeated using the average of the 5 values excluding outliers. If the average of all samples, excluding obvious visual outliers, was above the high risk cutoff the sample was presumed normal.

    b. If the average of the 3 tests were below the high risk cutoff a risk assessment was performed by MSPHL.

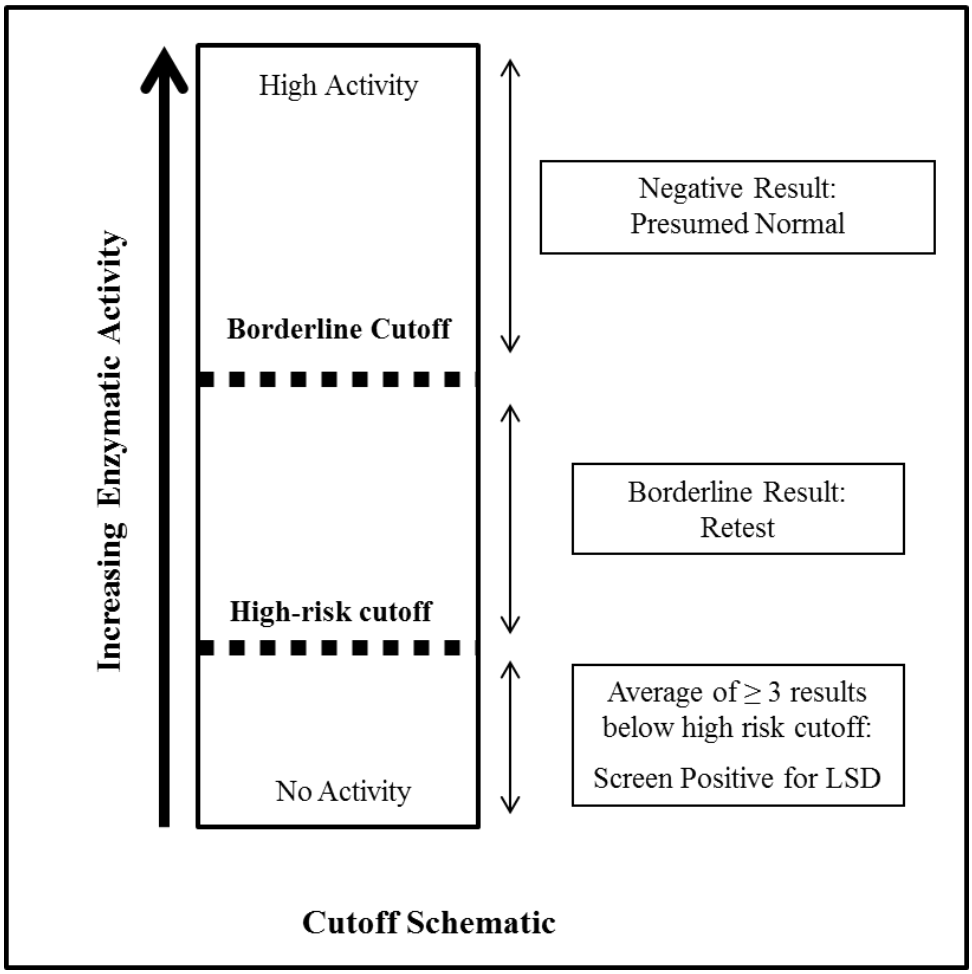


Figure 1: Schematic of borderline and high risk cutoffs with respect to enzyme activity and re-testing.

Risk Assessment

A number of criteria (described in the table below) were evaluated when assigning risk to babies that had ≥ 3 test results (the average of) below the high risk cutoff. Note that there was also some "experience-based judgment" by MSPHL personnel that was applied in the assessment.

Table 1: Risk Assessment Details

Risk Assessment Criteria	Description
*Additional Samples	When other test results for the enzyme activity were available for the same newborn and the other test results were in the presumed normal range, the risk level was reduced.
Age at Collection	In addition to the use of age specific cutoffs, MSPHL did not refer babies for LSD follow up based on samples collected at less than 24 hours of life since collection of a new DBS (re-screen) was mandated for those babies.
Sample Quality	If the samples were considered poor quality samples, they were generally not referred, since collection of a new DBS (re-screen) was mandated for those babies.
Transfusion Status	Transfused babies tend to have lower activity and MSPHL obtains at least one repeat screen; so they did not refer transfused newborns.

Risk Assessment Criteria	Description
Family History	If there is a known family member that is affected or a carrier for one of the diseases, then the risk would be significantly elevated.
Activity Value	The activity value in relationship to the cutoff was also considered; the lower the activity, the higher the risk level.
Transit Time / Birthplace	MSPHL utilized a courier system that transports > 95% of newborn samples from the birthing centers to the laboratory. Samples were mailed to MSPHL in the cases of home births or deliveries at birthing centers that do not participate in the courier program. In cases where the sample spent significant time in transit, the activity values were reduced. The risk level decreased with increased sample transit time.
Gestational Age	At least one repeat screen was required from babies that were born significantly premature (< 35 weeks gestational age). In cases where the birth was premature and another sample was expected, the assessment may have been postponed until the DBS from the repeat screen was received.
Age at Sample Collection	MSPHL used age at collection as a criterion in the risk assessment. For example, samples much greater than 14 days of life are expected to have decreased enzymatic activity values.
Other LSD results	MSPHL considered samples where an LSD enzyme was below the high risk cutoff and at least one other LSD enzyme was below the borderline cutoff to be a potentially poor quality sample; this dramatically reduced the risk level. Baebies stated that given the population distribution of all four enzymes and assuming that the activity values of the four enzymes are expected to be biochemically independent, Baebies estimated the likelihood of one assay below high risk and another below borderline was between 1 in 125,000 and 1 in 1,400,000. In addition to considering samples with an additional assay below borderline to be low risk, MSPHL also considered some DBS with additional LSD enzymes slightly above borderline (low-normal) to be low risk as well.
Other Test Results	MSPHL looked at results of the other newborn screening tests, but abnormal results for other tests did not significantly affect the risk level.
Other Altered Health Status	If the baby was indicated as “sick” (represented by a category of that designation on the screening card), MSPHL may have attempted to get more information about the illness and any effect that it may have on LSD results to consider its impact on the risk.

* Additional Samples: MSPHL evaluated all valid test results that they obtained for each newborn for the LSD enzymes for the risk analysis. This could include test results obtained from any previous screen (for babies that were screened more than once) and/or test results obtained from new punches from the same DBS and/or test results obtained from new punches from a different DBS from the same card (e,g., if a repeat screen was ordered for any reason, tests for all 4 disorders were run on the repeat screen).

If none of these conditions applied, and if the quality of the DBS was not considered poor owing to multiple low enzyme levels, then the baby was considered high risk and was referred to 1 of 4 contracted genetic referral centers for evaluation, confirmatory testing, and diagnosis.

Clinical Diagnosis

For babies determined to be high risk for the screened conditions and referred, true clinical status was determined by the methods summarized below:

Disorder/ Enzyme Analyzed	Confirmatory Tests	Possible Diagnoses	
		Affected	Normal
Mucopolysaccharidosis Type I (MPS I)/ IDUA	IDUA assay on leukocytes Mutation analysis	Attenuated Severe Genotype of unknown significance	Normal Carrier Pseudodeficiency
Pompe/ GAA	GAA assay on leukocytes Urine HEX4 assay Creatine kinase	Classical Infantile Onset Nonclassical Infantile Onset Late Onset Unknown Onset Genotype of Unknown Significance	Normal Carrier Pseudodeficiency
Gaucher/ GBA	GBA assay on leukocytes Mutation analysis	Neuronopathic Non-neuronopathic Unknown Onset Genotype of Unknown Significance	Normal Carrier
Fabry/ GLA	Male: GLA assay in leukocytes Mutation analysis Female: Mutation analysis	Classical Late Onset Genotype of Unknown Significance	Normal Pseudodeficiency

Newborn Screening Follow-Up Program

To calculate the false negative result rate for the assay, the clinical status of all newborns who tested negative would have to be known. However, it was impractical to perform confirmatory testing (or other suitable follow-up) on all negative babies. In order to collect information on potential false negative results during the clinical study, MSPHL worked in tandem with the metabolic centers. The metabolic centers that serve the state of Missouri collected surveillance information to identify any baby that was diagnosed with any of the screened conditions and participated in the clinical study (false negatives). Baebies stated that there were no reported cases of false negatives. There were 6 cases where parents refused follow up (out of 275 newborns referred for any one of the 4 conditions) and 3 newborns moved out of state. These babies were considered lost-to-follow up.

Study Phases

MSPHL conducted the study in several phases. MSPHL initially set the high risk cutoff values for each of the 4 LSDs by analyzing approximately 13,000 presumed normal de-identified DBS and 29 known affected DBS on the Baebies System. The cutoffs were chosen to ensure that all known affected samples would be detected

(sensitivity =100%) and to keep the potential false positive rate below 1% (specificity > 99%). The initial cutoffs were set at the levels described below:

Condition/ Enzyme	High Risk Cutoff (µmol/L/hour)	Borderline Cutoff (µmol/L/hour)
Mucopolysaccharidosis Type I (MPS I) / IDUA	4	5
Pompe/ GAA	8	10
Gaucher/ GBA	4.5	7
Fabry/ GLA	5.5	7

In the original submission, Baebies stated that the pivotal study ran from January 15, 2013 to January 14, 2015 and provided results for the clinical study throughout that time period. During the review of the submission, FDA had questions about multiple significant changes to the device during the pivotal study, and requested that Baebies provide information on how those changes may affect the way the results of the study should be interpreted. The changes included a significant change to the printed board of the cartridge (on August 16, 2013) and several changes to the assay protocol. In addition, the cutoffs were modified by MSPHL several times during the study (please refer to the section titled "changes to the cutoffs" provided for each assay below).

In response, Baebies proposed to retrospectively re-define the pivotal study period and divided the study into 2 phases. Baebies retrospectively defined a pilot phase; this phase was defined as the period when babies born between January 15, 2013 and August 26, 2013 were tested. Baebies retrospectively defined the pivotal clinical study as the period when babies born on or after August 27, 2013 were tested. This pivotal period lasted approximately 17 months (until January 14, 2015). During this (newly defined) pivotal period, the device was modified once. The modification, which occurred on November 17, 2014, was a formulation change to the stop buffer (the concentration of tween in the stop buffer was increased from 0.01% to 0.04%) intended to improve droplet movement and reduce the number of invalid test results (after the change, the invalid rate decreased from 7.56% to 5.97%)[a]. For reference, the original and modified study periods are as follows:

- Originally defined pivotal study period: January 15, 2013 to January 14, 2015
- Retrospectively Modified Study periods:
    - "Pilot" phase: January 15, 2013 to August 26, 2013
    - "Pivotal" phase: August 27, 2013 to January 14, 2015

[a] To support the change, DBS samples with known low enzymatic activity (i.e., samples with known screen positive results) were retested using the new formulation. These samples were either from newborns that were referred during the clinical study or diagnostic samples. The results were evaluated using the reformulated buffer against the high risk cutoff used at time of testing. All confirmed positive samples had results that fell below the high risk cutoff (i.e., low or no enzymatic activity; a screen positive result) for the respective assays.

Total Newborns Screened During the Study

In total (pilot and pivotal) 154,412 babies were screened (see table below):

	Total screened	# *excluded	# with no valid data	# included in analysis	# with 1 screen	# with 2 screens	# with 3 screens	# with 4+ screens
Total	154,412	701	14	153,697	136,309	13,579	3,249	560
Pilot	48,813	203	2	48,608	43,185	4,239	996	188
Pivotal	105,599	498	12	105,089	93,124	9,340	2,253	372

*All tests from a screen were excluded from the analysis for the following reasons:

- Screens with no recorded age
- If screens were collected when the baby was less than 24 hours old
- Screens with no valid test results
- If all screens from a baby were excluded, then the baby was excluded

Screens and Repeat Screens

For the entire study period, Baebies provided the following information about the number of babies with repeat screens and the reasons for the repeat screens:

- A total of 17,388 newborns with repeat screens during the total study period
- 6,052 newborns re-screened per NICU guidelines (gestational age < 34 wks, collection at < 24 hours of birth, birth weight < 2000 g, or the baby was transfused)
- 1,112 newborns born at 34 weeks were re-screened, likely due to premature status
- 10,224 newborns with re-screen collected for other reason (other screening results, poor quality samples, unknown)

Of note, repeat screens were not obtained as a result of abnormal LSD screening results during the study.

Overview Of Total Testing Performed During The Pivotal Phase of the Study

Baebies provided an overview of all testing performed during the pivotal phase of the study. A table that summarizes all testing performed (i.e., testing from the DBS card, retesting from the initial DBS card and all testing and repeat testing from all DBS cards obtained from subsequent screens) can be found in the Appendix.

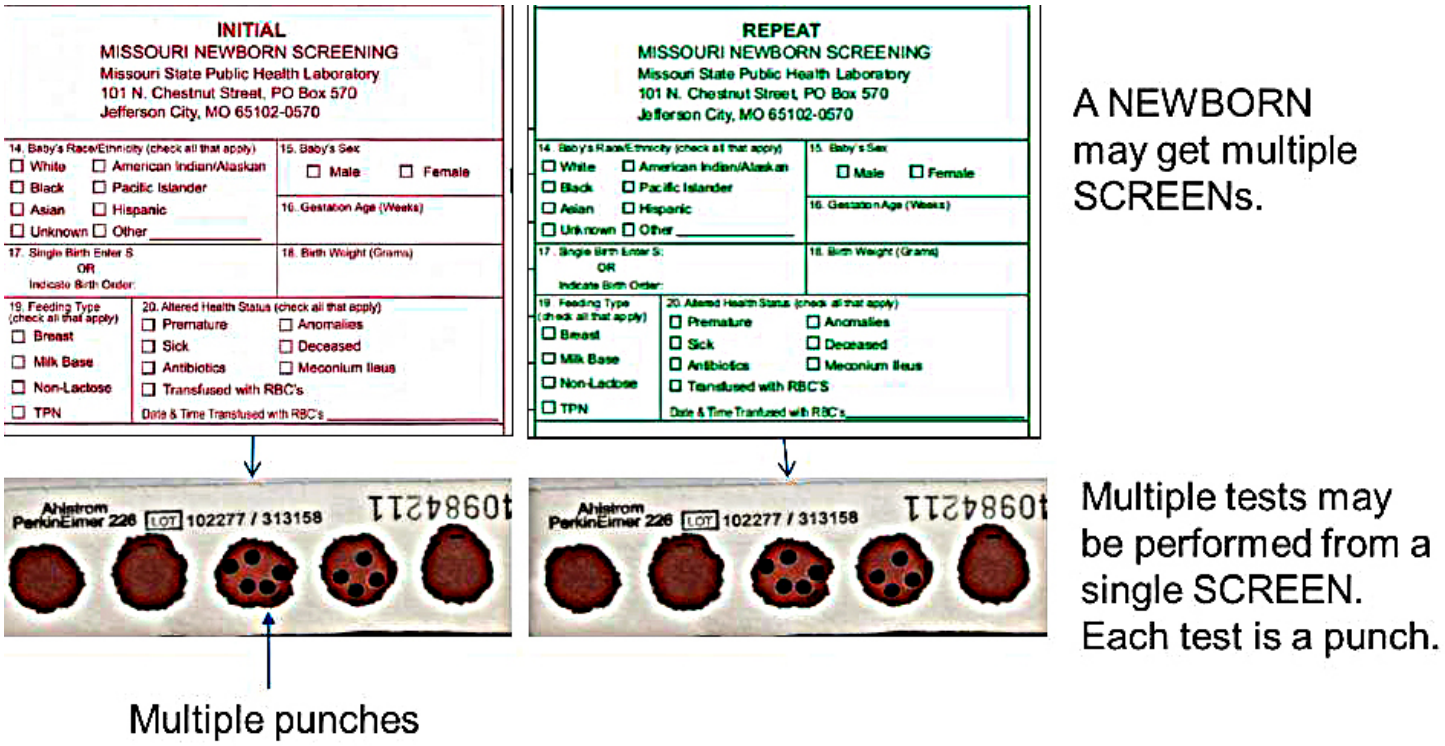
Assay Performance

In the following sections, we summarize the screening performance for each assay in the pivotal phase of the study as provided to us by Baebies. We frequently use the following terms which we define below for clarity:

- The term “screen” refers to the collection of a single blood collection card with multiple blood spots from a baby. Each baby can have a DBS card from the initial screen and one or more additional DBS cards collected during repeat screens.
- The term “test” refers to the assay result from a punch from a blood spot (each test uses one punch).

So that:

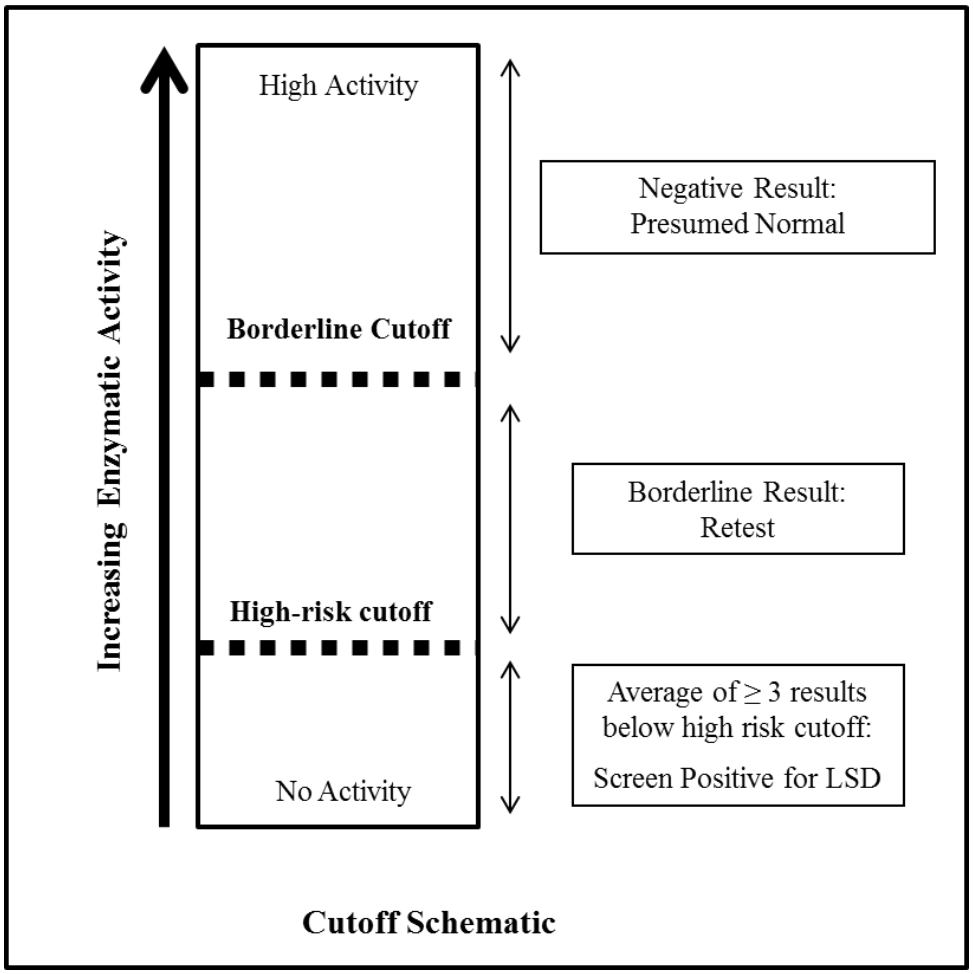
- One baby can have one DBS card from the initial screen and additional DBS cards from repeat screens (as needed).
- One baby can have an initial test and repeat tests from the DBS card from the initial screen and from the DBS card(s) from repeat screens (as needed).



(i.e., each DBS can be punched multiple times)

How Test Results Were Interpreted During the Study

For these disorders, affected babies have low enzyme activity. In order to screen the babies, 2 cutoffs were established; a high risk cutoff and a borderline cutoff. Babies with enzymatic activity below the high risk cutoff have a high likelihood of having the disorder while babies with enzymatic activity below the borderline cutoff were considered to have an undetermined risk. Babies with enzymatic activity above the borderline cutoff were determined to be at low risk for the screened condition and presumed normal. For convenience, here we briefly summarize how the test results were interpreted by MSPHL during the study. For a detailed description on the testing procedure, please refer to the section above titled “screening decision tree”.



1. All babies with enzymatic activity above the borderline cutoff were considered low risk for the screened condition and presumed normal.

2. Babies with enzymatic activity below the borderline cutoff were considered to have an undetermined risk and retested. If the average enzymatic activity (excluding outliers) of all tests was above the high risk cutoff, the baby was considered to have a low risk for the screened condition and presumed normal.

3. If the average enzymatic activity (excluding outliers) of all tests was below the high risk cutoff, the baby was considered to have a high risk of having the disorder and subjected to the risk assessment described above in table 1. If after the risk assessment the baby was considered to be at low risk for the screened condition, the baby was presumed normal. If the baby was still considered high risk for the screened condition after the risk assessment, the baby was referred for diagnostic follow-up.

4. Lower thresholds and upper thresholds for each enzyme were also evaluated to determine if the test results were valid (please see section titled “valid run and valid test results analysis” above).

B. Screening Results of the α-L-Iduronidase (IDUA) Enzyme Activity Assay to Screen for Mucopolysaccharidosis Type I (MPS-I) During the Pivotal Study Phase

We first present the results of the IDUA enzyme activity assay intended to screen babies for MPS I. For this study, two cutoffs – a high risk cutoff and a borderline cutoff – were used for all babies irrespective of age at the time the screen was collected. This is supported by the analysis of the reference interval (mostly based on the Quartile 1 (Q1), mean, median and Quartile 3 (Q3)) that Baebies calculated from the presumed normal babies following the pilot and pivotal. The IDUA enzyme activity does not appear to significantly change with age (see table titled "normal reference intervals for IDUA" below and for a detailed discussion on the age-related enzyme activity, please refer to the sponsor's Appendix A "Age At Collection Related Changes In Activity"). (Note: this is not true for the other three enzymes, which have age-specific cutoffs because the activity of those enzymes are known to change with age of the baby.)

Normal reference intervals for IDUA (calculated from the presumed normal babies)

		Quantile (µmol/L/hr)		Population Activities (µmol/L/hr)			
Age	n	0.10%	99.90%	Q1	Mean	Median	Q3
1-6 days	151,960	3.17	81.72	14.30	19.68	18.56	23.72
7-13 days	10,620	2.34	111.88	13.95	21.10	19.48	26.24
14+ days	12,880	1.82	90.97	13.82	21.90	19.99	27.55

The following cutoffs (in µmol/L/hr) were used during the clinical study:

IDUA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	4.0	4.0	4.0	5.0	5.0	5.0
2	5/15/2013	3.0	3.0	3.0	5.0	5.0	5.0
3*	7/3/2013	2.0	2.0	2.0	5.0	5.0	5.0
4	11/18/2014	1.5	1.5	1.5	5.0	5.0	5.0

* This change in cutoff occurred close to the end of the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

Baebies provided the following table summarizing the screening results for the IDUA assay during the pivotal phase of the study. This table summarized the results using the cutoffs that MSPHL was using at the time the samples were tested (i.e., though Missouri changed the cutoffs during the study, this data is analyzed using the cutoffs used at the time the newborn screen of each baby was performed).

		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,639	9,165	2,103	338	104,245
	At least one screen with first result below borderline	485	175	150	34	844
Average of all Tests	All screens w/ avg. above high risk	454	166	144	29	793
	At least one screen w/ avg. below high risk	31	9	6	5	51
	Referred	28	5	-	-	33
	Not Referred	3	4	6	5	18
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Hurler - Pseudodeficiency, False Positive	16	4			20
	Hurler - Normal, False Positive	8	1			9
	Hurler - Carrier, False Positive	2				2
	Hurler - True Positive					0
	Hurler - Refused	1				1
	Hurler - Moved	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		1	6	5	12
	Outliers Excluded	3				3
	Other Assay Below Borderline		2			2
	Later Sample From Newborn Above Cutoff		1			1
Performance Summary	Total Presumed Normal	93,096	9,335	2,253	372	105,056
	Total Presumed Affected	28	5	-	-	33
	True Positives	0	0	0	0	0
	False Positives	26	5	-	-	31
	False Positive Rate (FPR)	0.028%	0.054%	0.000%	0.000%	0.029%

For the IDUA assay the table above shows that overall 105,089 babies were included in the analysis during the retrospectively-defined pivotal study phase. Of these:

- 104,245 babies' initial test results were considered low risk for MPS-I and they were presumed normal.

- 844 babies' initial test results indicated an undetermined risk for MPS-I and they were subjected to additional testing.

Of these 844 babies, 793 babies' repeat test results indicated a low risk for MPS-I and they were presumed normal.
The remaining 51 babies had results that indicated that they were at high risk for MPS-I. These babies were subjected to the risk assessment described above in table 1. Based on the results of the risk assessment, of the 51 babies with high risk test results:

- 33 babies were still considered high risk for MPS-I and were referred to the metabolic centers for confirmatory diagnostic testing
- 18 babies were not referred because after the risk assessment they were not considered to be at high risk for MPS-I.

Of the 33 babies referred for diagnostic follow-up:

- 20 babies were determined to have a pseudodeficiency
- 2 babies were determined to be carriers
- 9 babies were determined to be normal
- 2 babies were not determined (1 refused diagnostic testing and 1 baby left the state and was lost to follow-up).

There were no (0) true positives detected during the pivotal study phase (as defined by Baebies post-hoc; see description on page 12 of the post-study division into pilot and pivotal phases). However, during the retrospectively-defined pilot phase of the study, one confirmed positive MPS-I baby was found.

Meanwhile for the 18 babies that were not referred the following factors were considered to reduce their risk of MPS-I:

- 12 babies had prior test results (from prior screens) that indicated that they were at low risk for MPS I. 1 baby had subsequent tests (from a subsequent screen) that indicated that the baby was at low risk for MPS I. As discussed above in the risk assessment section, MSPHL evaluated all available test results for babies when making the final risk determination for MPS I (and decision to either refer the baby or not to refer the baby). It is unclear if this same number of convenience-sample test results would be available to other laboratories/programs implementing this test system (if authorized for marketing).)
- 3 babies were not referred because MSPHL identified visual outliers among the test results. Once the visual outliers were excluded from the analysis, the average of the remaining test results indicated that the babies were at low risk for MPS I and they were presumed normal.
- 2 babies not referred because the other LSD assay results indicated high risk for those screened conditions. This was considered very unlikely, so the sample was presumed to be of low quality. Per MSPHL's risk analysis, this reduced the risk for MPS I and these babies were presumed normal.

A summary of the screening performance of the IDUA assay for all babies screened during the retrospectively-defined pivotal phase of the study is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for MPS I (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening Performance IDUA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,093	9,331	2,247	367	105,038
Test screen positive	31	9	6	5	51
Referred	28	5			33
Not referred	3	4	6	5	18
True positives	0	0	0	0	0

Estimated False Positive Rates

Baebies estimated the false positive rate of the screening procedure as 0.029% using the following formula:

The number of false positives (31) from the babies referred for diagnostic follow up (not including babies lost to follow-up)

divided by

The total number of babies screened (105,089) minus any confirmed positive babies (0) and minus any babies lost to follow up (2)

That is 31/105,087=0.029%

To calculate the false positive rate of the assay on the Seeker System, FDA used a different approach and the following formula:

> The number of babies with test results indicating high risk of MPS-I (i.e., below the "high-risk" cutoff) that did not have MPS I (51) minus the number of babies lost to follow-up (2)
>
> divided by
>
> The total number of newborn screened (105,089) minus the number of confirmed affected babies (0) minus any babies lost to follow up (2)
>
> That is 49/105,087=0.046%

The difference between the false positive rate calculated for MSPHL's screening procedure (0.029%) and the false positive rate of the assay calculated by FDA (0.046%) is that the false positive rate of the assay includes the babies with test results that indicated a high risk for MPS I that were not referred because the risk analysis lowered the risk for the baby. The reason for including these babies in the false positive rate for the device is that the risk analysis is a component of the laboratory's practice based on the expertise of the laboratory and is independent of the device. Manufacturers of newborn screening tests include false positive rate estimates for their device in their Instructions for Use so laboratories who install and use their system will know what to expect regarding this rate when implementing their testing. FDA would like panel input on whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or other formats of information to assist laboratories in implementing this test.

Estimated False Negative Rate

Based on the newborn screening follow-up program, Baebies reports no known false negatives for the IDUA assay since there have been no babies that were screened during the study period referred to the metabolic centers and diagnosed with MPS I. FDA would like panel input on whether this type of analysis is adequate to conclude that the IDUA assay had no false negatives. Note that in a retrospective analysis of confirmed positive samples that included the initial screening and repeat testing of the samples as well as additional testing of the samples for quality assurance purposes, a sample from 1 confirmed MPS I baby that was tested 13 times over the study period was evaluated. In this small study to evaluate analytical false negatives, all 13 tests of the confirmed MPS I baby were always correctly categorized as high risk for MPS I.

Information on the Confirmed MPS I Baby

During the retrospectively-defined pilot phase of the clinical study, one newborn was identified by the Baebies Seeker System and diagnosed with MPS I. The one confirmed MPS I baby had 2 valid screens. The baby was referred for diagnostic follow-up for MPS I based on the results of the first screen. The reason for the second screen was not provided to Baebies.

Modifications to the Cutoffs During the Clinical Study

The IDUA cutoffs were modified several times by MSPHL during the pilot phase of the study and once during the pivotal phase. Below are the different cutoffs (in µmol/L/hr) used during the entire clinical study:

IDUA cutoffs		High Risk			Borderline Risk		
Cutoff period	Date of cutoff change	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	4.0	4.0	4.0	5.0	5.0	5.0
2	5/15/2013	3.0	3.0	3.0	5.0	5.0	5.0
3*	7/3/2013	2.0	2.0	2.0	5.0	5.0	5.0
4	11/18/2014	1.5	1.5	1.5	5.0	5.0	5.0

* This change in cutoff occurred close to the end of the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

The timing, reason and a description of the change provided by MSPHL to the IDUA cutoffs is summarized in the table below:

Date of Change	Cutoff in use	Summary of information reviewed	Cutoff decision
05/15/2013	4.0µmol/L/hr	Lab's Referral rate – 0.083% *Lab's False positive rate – 0.079% Referral rate was high for the expected incidence and false positive rate was high. One true positive MPS I specimen identified with average activity - 1.40 µmol/L/hr which is well below the cutoff used.	Decrease from 4.0 to 3.0 µmol/L/hr
07/03/2013	3.0µmol/L/hr	Lab's Referral rate – 0.093% *Lab's False positive rate – 0.086% Referral rate and false positive rate did not decrease. One true positive MPS I specimen (repeat specimen from previously identified newborn) identified with average activity 0.20 µmol/L/hr which is well below the cutoff used.	Decrease from 3.0 to 2.0 µmol/L/hr
11/18/2014	2.0µmol/L/hr	Lab's Referral rate – 0.030% *Lab's False positive rate – 0.030% Referral rate and false positive rate decreased considerably. Cutoff still above the affected activity from previous cutoff periods.	Decrease from 2.0 to 1.5 µmol/L/hr

* This false positive rate describes the false positive rate of MSPHL's screening procedure and not the false positive rate for the device as FDA has calculated it (please refer above to the section "estimated false positive rates" for a discussion on how these differ).

To explore the impact of the changes to the cutoffs during the pivotal study, Baebies provided a table summarizing the screening performance of the device for the entire pivotal phase of the study applying the final cutoffs used by MSPHL. The screening results are summarized in the table below:

		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,639	9,165	2,103	338	104,245
	At least one screen with first result below borderline	485	175	150	34	844
Average of all Tests	All screens w/ avg. above high risk	459	169	145	32	805
	At least one screen w/ avg. below high risk	26	6	5	2	39
	Referred	25	4	-	-	29
	Not Referred	1	2	5	2	10
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Hurler - Pseudodeficiency, False Positive	15	3			18
	Hurler - Normal, False Positive	6	1			7
	Hurler - Carrier, False Positive	2				2
	Hurler - Refused	1				1
	Hurler - Moved	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		1	5	2	8
	Outliers Excluded	1				1
	Other Assay Below Borderline		1			1
	Later Sample From Newborn Above Cutoff					0
Performance Summary	Total Presumed Normal	93,099	9,336	2,253	372	105,060
	Total Presumed Affected	25	4	-	-	29
	True Positives	0	0	0	0	0
	False Positives	23	4	-	-	27
	False Positive Rate (FPR)	0.025%	0.043%	0.000%	0.000%	0.026%

The impact of the changes to the cutoffs is that 12 babies that were considered high risk for MPS I (based on the cutoffs in use by MSPHL at the time these babies were screened) would no longer be considered high risk (i.e., during routine screening, 51 babies were considered high risk based on the test result and using the final cutoffs, 39 babies would be considered high risk). The false positive rate for MSPHL's screening procedure goes down to 0.026% (compared to 0.029%) and the false positive rate of the test (i.e., FDA's analysis) also goes down to 0.035% compared to 0.046%. There were no babies diagnosed with MPS I during the pivotal phase of the study. The test results of the baby that was diagnosed with MPS I during the pilot phase of the study would still be considered high risk for MPS I using the final cutoffs (and all cutoffs using during the clinical study).

A summary of the screening performance of the IDUA assay for all babies screened during the retrospectively-defined pivotal phase of the study using the final cutoffs is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for MPS I (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance IDUA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,098	9,334	2,248	370	105,050
Test screen positive	26	6	5	2	39
Referred	25	4			29
Not referred	1	2	5	2	10
True positives	0	0	0	0	0

FDA would like the panel's input on whether the analysis of the clinical study should use the cutoffs used when testing each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff – cutoff period 4). This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

Analytical Performance of the IDUA Assay

Quantitative assays are designed to report patient results within a measurement range that is analytically reliable (e.g., precise, reproducible, accurate). The reliability of the assay is particularly important around any clinical decision cutoff. During the review of the data provided in support of the IDUA assay, FDA noted the following.

Precision: Baebies performed a study to estimate the imprecision of the IDUA assay using 4 analyzers and 3 reagent lots, and performed testing during 21 non-consecutive days, with 2 runs per day and 2 dried blood spot (DBS) punches of each specimen per run. Baebies tested a total of 336 replicate results for each specimen. In this study 11 samples were identified as invalid and 3 IDUA test results were identified as "high statistical outliers". The results of the study are summarized below (please note that the outliers are included in this analysis). In the following table the repeatability (or within-run precision) estimate includes the instrument run as the component (or condition) of variability and the reproducibility estimate includes the instrument run, reagent lot, instrument and day as the components (or conditions) of variability. As a reminder, the final cutoffs during the pivotal study were 1.5 and 5.0 µmol/L/h.

Mean µmol/L/h	N	Repeatability (%CV)	Between Lot (%CV)	Between Instrument (%CV)	Between Day (%CV)	Reproducibility (%CV)
2.40	331	74.6%	22.9%	0%	34.6%	82.1%
3.53	334	22.7%	15.6%	0%	7.4%	27.2%
6.22	335	26.4%	11.1%	0%	0.0%	28.5%
12.09	334	15.4%	11.0%	0%	0.0%	18.8%
24.06	335	9.0%	9.8%	0%	0.0%	14.2%

The coefficient of variability (CV) was calculated for each imprecision estimate. The CV is calculated by dividing the standard deviation by the mean of the sample and describes the extent of variability in relation to the mean of the sample. The highlighted sections in the table above denote the imprecision of the IDUA assay near the cutoffs used in the pivotal study.

Detection Limits: Baebies evaluated the detection limits of this assay following a recognized guideline[10] using 3 lots of their reagent and estimated the following detection limits:

- The limit of the blank (LoB) was defined as the highest analyte concentrations expected to be found when replicates of a sample containing no analyte are tested with 95% confidence. This is often a way of determining what concentration(s) the assay cannot distinguish from "noise." Baebies estimates that the LoB of the IDUA assay was 1.78 µmol/L/h (note the final high-risk cutoff for the assay was 1.5 µmol/L/h).

- The limit of detection (LoD) was defined as the lowest analyte concentration likely to be reliably distinguished from a blank sample with 95% confidence. The LoD of the IDUA assay was determined to be 2.77 µmol/L/h.

- The limit of quantification was defined as the lowest concentration where the total imprecision was ≤ 1.5 µmol/L/h or 20% CV whichever was greater. Baebies estimated that the LoQ for the IDUA assay was 2.77 µmol/L/h which was the concentration where the imprecision was less than 1.5 µmol/L/h (and the CV could be as high as 54%). Based on the data provided in support of the LoQ of the assay, FDA estimates that the LoQ based on an imprecision goal of 20% CV (which is the typical imprecision goal

for the LoQ of quantitative assays) is approximately 3.7 µmol/L/h (although FDA notes that this estimate is not consistent with the precision evaluation of the test which demonstrated higher imprecision at this concentration).

Outliers: Baebies identified statistical outliers in their analytical studies in support of the IDUA assay (and all assays). In fact, MSPHL also identified "visual outliers" during the clinical study for the IDUA assay when clinical samples needed to be tested in multiple replicates. Removing visual outliers from the analysis resulted in a change to the risk for MPS I for 3 babies that were initially considered high risk for MPS I. In the analytical studies (precision, detection limits) these outliers impacted the performance of the assay.

Stability of the IDUA samples (transport stability): Baebies performed a study showing the impact of shipping the DBS for up to 5 days at the following ambient conditions: 10ºC and 20% relative humidity (RH), 10ºC and 80% RH, 45ºC and 20% RH, 45ºC and 80% RH and 25ºC and 50% RH. The impact on the test result is summarized in the table below:

Average IDUA enzymatic activity for DBS samples before and after simulated transport

Condition	Day	Sample 3 Concentration µmol/L/hr		Sample 5 Concentration µmol/L/hr		Sample 10 Concentration µmol/L/hr		Sample 15 Concentration µmol/L/hr	
		before	after	before	after	before	after	before	after
10C, 20%RH	1	3.8	3.6	4.0	4.5	11.5	12.8	25.8	26.0
	3	3.8	4.8	4.3	4.5	12.0	13.7	27.2	28.9
	5	4.4	4.7	5.0	4.8	12.3	12.6	27.2	29.1
10C, 80%RH	1	3.8	3.8	4.2	4.6	11.9	12.6	26.1	27.9
	3	3.6	4.3	4.0	4.1	11.0	12.5	24.7	26.0
	5	3.8	4.0	4.2	3.9	12.1	13.7	26.6	28.0
25C, 50%RH	1	4.5	4.5	4.7	5.4	13.0	13.5	26.6	26.0
	3	4.6	4.8	5.4	6.0	13.2	13.3	27.6	27.3
	5	3.8	3.7	4.5	5.0	12.3	12.9	25.7	25.7
45C, 20%RH	1	5.5	4.3	4.4	5.3	12.6	12.4	27.2	22.9
	3	4.0	4.3	4.5	4.8	11.8	10.4	24.9	18.7
	5	4.4	4.8	4.8	5.5	12.7	11.4	27.2	18.5
45C, 80%RH	1	4.2	3.9	4.5	4.5	12.4	9.9	26.9	17.7
	3	3.9	3.2	4.3	3.4	11.6	6.2	24.8	6.7
	5	4.0	3.6	4.5	4.1	12.2	5.0	26.1	5.0

The samples above presented in red font are outside the acceptance criteria for acceptable stability defined by the sponsor (i.e., recovery values <85% and >115% for enzyme concentrations greater than ≈6 µmol/L/hr and mean difference in enzymatic activity outside +/-1 µmol/L/h for enzyme concentrations below ≈6 µmol/L/hr were considered to be significantly impacted). The highlighted fields show samples with targeted concentrations set to represent normal and high normal samples that upon transport for 5 days at 45ºC and 80% relative humidity fall at the borderline cutoff or very close to the borderline cutoff (which is 5 µmol/L/hr for this assay). These shipping conditions might be experienced in certain regions of the United States, such as in the southern states especially, if shipped at ambient conditions.

This study demonstrates that the enzyme activity will decrease when exposed to high temperature and/or humidity, and should be taken into account in certain geographic areas/seasons. This degradation could lead to

false positive results. FDA is asking for panel input on whether the panel is aware of any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity as result of standard shipping conditions.

Analytical Performance of the Assay at the Cutoffs

In summary:

- The high risk cutoffs used during the pivotal study ranged from 1.5 to 2 µmol/L/h and the assay at this concentration range has an estimated repeatability greater than 74.6% CV (and estimated reproducibility greater than 82.1% CV). However, when the high risk cutoff is used, there were always multiple test results from the same DBS (any test result from previous screens or subsequent screens) available from each baby for analysis and interpretation.

- The borderline cutoff was 5 µmol/L/h and at this concentration the assay has an estimated repeatability between 22.7 and 26.4%CV (and estimated reproducibility between 27.2 and 28.5% CV).

- The high risk cutoffs used during the pivotal study (see table below) were set below the LoD (i.e., 2.77 µmol/L/h) and the cutoffs used in period 2 were set below the LoB of the assay.

IDUA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
Pivotal 1	7/3/2013	2.0	2.0	2.0	5.0	5.0	5.0
Pivotal 2	11/18/2014	1.5	1.5	1.5	5.0	5.0	5.0

  The high risk cutoffs were set below where the assay can reliably detect the analyte; the second cutoff used was below the limit of blank (i.e., a range where more than 95% of samples with no analyte return results). Analytically, both sets of cutoffs used in these studies are indistinguishable from each other and from the LoD of the assay (i.e., 2.77µmol/L/h). Cutoffs for most quantitative assays used in clinical practice are typically set in regions where the analyte can be reliably measured (i.e., above the LoQ of the assay defined by a clinically acceptable performance goal).

- There were more outliers than is typical for assays of this type (statistical outliers during the analytical studies and visual outliers during the clinical study).

- Transport at certain ambient conditions resulted in decreased enzyme activity. We are aware that laboratories have reported seasonal variation with other analyte levels (i.e., galactosemia assays[b]).

To conclude, for this clinical use, newborn screening test manufacturers typically provide clinical data in the package insert that is analyzed with cutoffs not only set so that no known positive babies are incorrectly categorized (i.e., presumed normal instead of presumed positive for the screened condition and vice versa) but the cutoffs are also analytically valid (e.g., precise, reproducible). The rationale for this is so that laboratories can clearly understand the clinical performance of a screening assay set to not miss any known true positive babies with analytically reliable cutoffs.

[b] http://health.mo.gov/living/families/genetics/newbornscreening/pdf/newbornscreeningreport2013.pdf

To summarize FDA's questions for the panel, we are seeking input on:

- Whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or another specified analysis?

- Whether the panel has a recommendation on how to estimate the false negative rate of this device?

- Whether the analysis of the clinical study should use the cutoffs used to test each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff)? This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

- Whether performance characteristics for age-specific cutoffs, when used, should be provided in the Instructions for use?

- Whether there are any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity standard shipping conditions, including high temperature and humidity?

- Whether the analytical performance of the assays at the cutoffs (e.g., precision, detection limits, outliers, performance of confirmed positive samples upon retesting) is sufficient for safe and effective use (i.e., does the panel consider that Baebies has demonstrated adequate analytical validity of the assays)?

C. Screening Results of the α-D-glucosidase (GAA) Assay to Screen for Pompe During the Pivotal Study Phase

For the pivotal study, two cutoffs – a high risk cutoff and a borderline cutoff – were used to screen for Pompe. In contrast to the IDUA assay, the high-risk cutoffs used were different depending on the age of the baby at the time the screen was collected. These age specific cutoffs were developed to account for the change in enzyme activity (see table titled "normal reference intervals for GAA" below) as a function of the baby's age at the time of sample collection. However, these age-specific cutoffs were added during the pilot and pivotal phases of the study. MSPHL added a different cutoff for newborns ≥ 14 days of age in May 16, 2013 and a different cutoff for newborns 7-13 days of age on June 23, 2014 as the laboratory noted these changes in the enzyme levels due to age of the newborn when the specimen was collected. The retrospective analysis of the reference interval (mostly based on the Quartile 1 (Q1), mean, median and Quartile 3 (Q3) calculated from the presumed normal babies screened during the entire clinical study (pilot and pivotal)), confirmed that the GAA enzyme activity significantly changes with age (for a detailed discussion on the changes, please refer to Appendix A provided by Baebies "Age At Collection Related Changes In Activity").

Reference intervals for GAA (calculated from the presumed normal babies)

		Quantile (μmol/L/hr)		Population Activities (μmol/L/hr)			
Age	n	0.10%	99.90%	Q1	Mean	Median	Q3
1-6 days	151,960	7.13	81.72	20.05	27.00	25.42	32.15
7-13 days	10,620	5.23	111.88	16.55	23.52	21.19	27.48
14+ days	12,880	4.27	90.97	13.56	19.29	17.48	22.67

The following cutoffs were used during clinical study:

GAA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	8.0	8.0	8.0	10.0	10.0	10.0
2	3/9/2013	7.0	7.0	7.0	10.0	10.0	10.0
3*	5/16/2013	7.0	7.0	4.5	10.0	10.0	10.0
4	11/12/2013	7.2	7.2	4.5	10.0	10.0	10.0
5	6/23/2014	7.2	4.5	4.5	10.0	10.0	10.0

* This change in cutoff occurred in the middle of the retrospectively-defined pilot phase; this cutoff was used at the beginning of the retrospectively-defined pivotal phase (defined as starting on August 27, 2013).

Baebies provided the following table summarizing the screening results for the GAA assay during the pivotal phase of the study. This table summarized the results using the cutoffs that MSPHL was using at the time the samples were tested (i.e., though Missouri changed the cutoffs during the study, this data is analyzed using the cutoffs used at the time the newborn screen of each baby was performed).

		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,462	8,938	1,980	311	103,691
	At least one screen with first result below borderline	662	402	273	61	1,398
Average of all Tests	All screens w/ avg. above high risk	607	369	257	55	1,288
	At least one screen w/ avg. below high risk	55	33	16	6	110
	Referred	32	8	4	1	45
	Not Referred	23	25	12	5	65
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Pompe - Normal, False Positive	15	4	3	1	23
	Pompe - True Positive	6	1			7
	Pompe - Moved					0
	Pompe - Carrier, False Positive	6	1			7
	Pompe - Pseudodeficiency, False Positive	5	2	1		8
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		17	9	4	30
	Later Sample From Newborn Above Cutoff	1	7	3	1	12
	Other Assay Below Borderline	11	1			12
	Outliers Excluded	8				8
	Transfused	2				2
	Multiple Reasons	1				1
Performance Summary	Total Presumed Normal	93,092	9,332	2,249	371	105,044
	Total Presumed Affected	32	8	4	1	45
	True Positives	6	1	0	0	7
	False Positives	26	7	4	1	38
	False Positive Rate (FPR)	0.028%	0.075%	0.178%	0.269%	0.036%

For the GAA assay the table above shows that overall 105,089 babies were included in the analysis during the retrospectively-defined pivotal study phase. Of these:

- 103,691 babies' initial test results were considered low risk for Pompe and they were presumed normal.
- 1,398 babies' initial test results indicated an undetermined risk for Pompe and they were subjected to additional testing.

Of the 1,398 babies subjected to additional testing, 1,288 babies' repeat test results indicated a low risk for Pompe and they were presumed normal.

This left 110 babies that had results that indicated that they were at high risk for Pompe. These babies were subjected to the risk assessment described above in table 1. Based on the results of the risk assessment, 45 of the 110 babies were still considered high risk for Pompe and were referred to the metabolic centers for confirmatory diagnostic testing while 65 babies were not referred because after the risk assessment they were not considered to be at high risk for Pompe.

Of the 45 babies that were referred for diagnostic follow-up:

- 7 were diagnosed with Pompe,
- 8 were determined to have a pseudodeficiency,
- 7 were determined to be carriers, and
- 23 were determined to be normal.

Meanwhile for the 65 babies that had high risk test results but were not referred following the risk analysis, the following factors were considered to reduce their risk of Pompe:

- 30 babies had prior test results (from prior screens) that indicated that they were at low risk for Pompe. 12 babies had subsequent tests (from a subsequent screen) that indicated that the baby was at low risk for Pompe. (As discussed above in the risk assessment section and in the IDUA section, during the risk assessment MSPHL evaluated all test results from all tests and retests from all screens and re-screens available for the baby to make their final risk determination. Again, it is unclear if this same number of convenience-sample test results would be available to other laboratories/programs implementing this test system.
- 8 babies were not referred because MSPHL identified visual outliers among the test results. Once the visual outliers were excluded from the analysis, the average of the remaining test results indicated that the babies were at low risk for Pompe and they were presumed normal.
- 12 babies were not referred because the other LSD assay results indicated high risk for those screened conditions. This was considered very unlikely and the sample was presumed to be of low quality. Per MSPHL's risk analysis, this reduced the risk for Pompe and these babies were presumed normal.
- 2 babies were not referred because they were transfused
- 1 baby was not referred for "multiple reasons."

A summary of the screening performance of the GAA assay for all babies screened during the retrospectively-defined pivotal phase of the study is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Pompe (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GAA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,069	9,307	2,237	366	104,979
Test screen positive	55	33	16	6	110
Referred	32	8	4	1	45
Not referred	23	25	12	5	65
True positives	6	1	0	0	7

Estimated False Positive Rates

Baebies estimated the false positive rate of MSPHL's screening procedure as 0.036% using the following formula:

> The number of false positives (38) from the babies referred for diagnostic follow up (not including babies lost to follow-up)
>
> divided by
>
> The total number of babies screened (105,089) minus any confirmed positive babies (7) and minus any babies lost to follow up (0)
>
> That is 38/105,082=0.036%

To calculate the false positive rate of the assay on the Seeker System, FDA used a different approach and the following formula:

> The number of babies with test results indicating high risk of Pompe (i.e., below the high risk cutoff) that did not have Pompe (103) minus the number of babies lost to follow-up (0)
>
> divided by
>
> The total number of newborn screened (105,089) minus the number of confirmed affected babies (7) minus any babies lost to follow up (0)
>
> That is 103/105,082=0.098%

As discussed above for IDUA, the difference between the false positive rate calculated for MSPHL's screening procedure (0.036%) and the false positive rate of the assay calculated by FDA (0.098%) is that the false positive rate of the assay includes the babies with test results that indicated a high risk for Pompe that were not referred because the risk analysis lowered the risk for the baby. The reason for including these babies in the false positive rate for the device is that the risk analysis is a component of the laboratory's practice based on the expertise of the laboratory and is independent of the device. Manufacturers of newborn screening tests include false positive rate estimates for their device in their Instructions for Use so laboratories who install and use their system will know what to expect regarding this rate when implementing their testing. FDA would like panel input on whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or other formats of information to assist laboratories in implementing this test.

Estimated False Negative Rate

Based on the newborn screening follow-up program, Baebies reports no known false negatives for the GAA assay since there have been no babies that were screened during the study period referred to the metabolic centers and diagnosed with Pompe.

To provide additional information on the false negative rate, Baebies performed a retrospective analysis of the initial screening and repeat testing of confirmed positive Pompe samples as well as additional testing of the samples for quality assurance purposes. In this analysis, 139 tests from several confirmed positive Pompe samples tested during the course of the study were evaluated. The results of this analysis are provided below:

Table 4 - Results Above Cutoff Based on Cutoff Used – GAA – all samples > 24h

Cutoff	# of Tests of Affected Samples	# of Tests above cutoff	% of tests above cutoff
High Risk (at time of test)	139	12	8.6%
Borderline (at time of test)	139	0	0.0%
High Risk (final)	139	16	11.5%
Borderline (final)	139	0	0.0%

The result of each test was evaluated based on the cutoffs that were in use when the sample was tested (and described above as "at the time of test") and the cutoffs that were in use at the end of the study (and described above as "final"). No test results were above the borderline cutoff (presumed normal), however 8.6% fell above the high risk cutoff (in the borderline region between high risk and borderline cutoff) used at the time the sample was tested and 11.5% fell above the final high risk cutoff used at the end of the study. There was no root cause analysis given for the analytical false negatives detected.

FDA would like panel input on whether this type of analysis is adequate to conclude that the GAA assay has no false negatives.

Information on Confirmed Positive Babies

During the pivotal clinical study, 7 newborns were diagnosed with Pompe. One confirmed positive Pompe baby had multiple screens. This neonate was identified and referred following the first valid screen. The affected newborn was detected on the first screen, however the additional screen (for other reasons) also had positive results.

Modifications to the Cutoffs During the Clinical Study

The GAA cutoffs were modified several times by MSPHL during the retrospectively-defined pilot phase of the study and twice during the retrospectively-defined pivotal phase. Below are the different cutoffs (in µmol/L/hr) used during the entire clinical study:

GAA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	8.0	8.0	8.0	10.0	10.0	10.0
2	3/9/2013	7.0	7.0	7.0	10.0	10.0	10.0
3*	5/16/2013	7.0	7.0	4.5	10.0	10.0	10.0
4	11/12/2013	7.2	7.2	4.5	10.0	10.0	10.0
5	6/23/2014	7.2	4.5	4.5	10.0	10.0	10.0

* This change in cutoff occurred before the end of the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

The timing, reason and a description of the change provided by MSPHL to the GAA cutoffs is summarized in the table below:

Date of cutoff change	Cutoff in use	Summary of information reviewed	Cutoff decision
03/09/2013	8.0µmol/L/hr	Lab's Referral rate – 0.060% *Lab's False positive rate – 0.038% 3 true positive Pompe specimens with average activity 4.11, 5.22, 3.33 µmol/L/hr Referral rate was considered to be high given prior incidence information. Difference between cutoff (8.0 µmol/L/hr) and highest affected (5.22 µmol/L/hr) was large enough to consider a lowering of the cutoff.	All cutoffs decrease from 8.0 to 7.0 µmol/L/hr
5/16/2013	7.0µmol/L/hr	The distribution of the 14+ day babies was evaluated and a different cutoff was introduced. The high risk 14+ cutoff was set to the same percentile as the cutoff percentile (i.e., 0.07^{th}) for the 1-6 day babies.	14+ day high risk cutoff introduced and reduced from 7.0 to 4.5 µmol/L/hr
11/12/2013	7.0µmol/L/hr	Lab's Referral rate – 0.049% *Lab's False positive rate – 0.032% 8 true positive Pompe specimens identified with average activity - 4.41, 6.46, 6.63, 3.90, 5.16, 6.56, 6.81, 4.38, and 6.10 µmol/L/hr. One Pompe positive specimen (6.81 µmol/L/hr) very close to cutoff (7.0 µmol/L/hr) prompting an increase in the cutoff.	Increase 1-6 day and 7-13 day high risk cutoff from 7.0 to 7.2 µmol/L/hr
6/23/2014	7.2 µmol/L/hr	The distribution of the 7-13 day babies was evaluated and a different cutoff was introduced. The high risk 7-13 day cutoff was set to the same percentile as the cutoff percentile for the 1-6 day babies.	7-13 day high risk cutoff introduced and reduced from 7.2 to 4.5 µmol/L/hr

* This false positive rate describes the false positive rate of MSPHL's screening procedure and not the false positive rate for the device as FDA has calculated it (please refer above to the section "estimated false positive rates" for a discussion on how these differ).

To explore the impact of the changes to the cutoffs during the pivotal study, Baebies provided a table summarizing the screening performance of the device for the entire pivotal phase of the study applying the final cutoffs used by MSPHL. The screening results are summarized in the table below:

GAA - Pivotal Phase						
		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,462	8,938	1,980	311	103,691
	At least one screen with first result below borderline	662	402	273	61	1,398
Average of all Tests	All screens w/ avg. above high risk	607	373	259	55	1,294
	At least one screen w/ avg. below high risk	55	29	14	6	104
	Referred	32	8	3	1	44
	Not Referred	23	21	11	5	60
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Pompe - Normal, False Positive	15	4	2	1	22
	Pompe - True Positive	6	1			7
	Pompe - Carrier, False Positive	6	1			7
	Pompe - Pseudodeficiency, False Positive	5	2	1		8
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		13	8	4	25
	Later Sample From Newborn Above Cutoff	1	7	3	1	12
	Other Assay Below Borderline	12	1			13
	Outliers Excluded	8				8
	Transfused	1				1
	Multiple Reasons	1				1
Performance Summary	Total Presumed Normal	93,092	9,332	2,250	371	105,045
	Total Presumed Affected	32	8	3	1	44
	True Positives	6	1	0	0	7
	False Positives	26	7	3	1	37
	False Positive Rate (FPR)	0.028%	0.075%	0.133%	0.269%	0.035%

The impact of the changes to the cutoffs is that 6 babies that were considered high risk for Pompe (based on the cutoffs in use by MSPHL at the time these babies were screened) would no longer be considered high risk (i.e., during routine screening, 110 babies were considered high risk for Pompe based on the test result and using the final cutoffs, 104 babies would be considered high risk for Pompe). The false positive rate for MSPHL's screening procedure is consistent at 0.035% (compared to 0.036%) and the false positive rate of the test (i.e., FDA's analysis) also remains consistent at 0.092% using the final cutoffs compared to 0.098% based on the cutoffs used at testing. All confirmed positive Pompe babies (including those diagnosed during the pilot phase of the study) would still be considered high risk for Pompe if the final cutoffs (and all cutoffs used during the study) are applied to the original test results.

A summary of the screening performance of the GAA assay for all babies screened during the retrospectively-defined pivotal phase of the study, using the final cutoffs is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Pompe (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GAA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,069	9,311	2,239	366	104,985
Test screen positive	55	29	14	6	104
Referred	32	8	3	1	44
Not referred	23	21	11	5	60
True Positives	6	1	0	0	7

FDA would like the panel's input on whether the analysis of the clinical study should use the cutoffs used when testing each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final

cutoff – cutoff period 5). This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

Analytical Performance of the GAA Assay

During our review of the analytical validation data provided in support of the GAA assay we noted the following:

Precision: Baebies performed a study to estimate the imprecision of the GAA assay using 4 analyzers and 3 reagent lots and performed testing during 21 non-consecutive days, with 2 runs per day and 2 dried blood spot (DBS) punches of each specimen per run. Baebies tested a total of 336 replicate results for each specimen. In this study 11 samples were identified as invalid. The results of the precision study are summarized in the table below. In the following table the repeatability (or within-run precision) estimate includes the instrument run as the component (or condition) of variability and the reproducibility estimate includes the instrument run, reagent lot, instrument and day as the components (or conditions) of variability. As a reminder, the final cutoffs during the pivotal study were 7.2 and 4.5 (high risk cutoffs) and 10.0 (borderline cutoff) µmol/L/h.

Mean µmol/L/h	N	Repeatability (%CV)	Between Lot (%CV)	Between Instrument (%CV)	Between Day (%CV)	Reproducibility (%CV)
4.29	331	15.6%	5.8%	0%	14%	17.0%
6.27	334	15.2%	4.1%	0%	0.0%	15.8%
9.59	335	9.9%	7.0%	0%	0.0%	12.0%
18.06	334	13.6%	5.9%	0%	0.0%	14.8%
27.37	335	11.3%	6.3%	1.6%	0.0%	12.9%

The coefficient of variability (CV) was calculated for each imprecision estimate. The CV is calculated by dividing the standard deviation by the mean of the sample and describes the extent of variability in relation to the mean of the sample. The highlighted sections in the table above denote the imprecision of the GAA assay near the cutoffs (borderline and high risk) used in the pivotal study.

Detection Limits: Baebies evaluated the detection limits of this assay following a recognized guideline[10] using 3 lots of their reagent and estimated the following detection limits:

- The limit of the blank (LoB) was defined as the highest analyte concentration expected to be found when replicates of a sample containing no analyte are tested with 95% confidence. This is often a way of determining what concentration(s) the assay cannot distinguish from "noise." Baebies estimated that the LoB of the GAA assay was 0.50 µmol/L/h.

- The limit of detection (LoD) was defined as the lowest analyte concentration likely to be reliably distinguished from a blank sample with 95% confidence. The LoD of the GAA assay was determined to be 2.18 µmol/L/h. However, FDA notes that including statistical outliers, the LoD could be as high as 5.36 µmol/L/h (note one of the high-risk cutoffs for the assay is 4.5 µmol/L/h).

- The limit of quantification was defined as the lowest concentration where the total imprecision was ≤ 1.5 µmol/L/h or 20% CV whichever was greater. Baebies estimated that the LoQ for the GAA assay was 2.18 µmol/L/h which was the concentration where the imprecision was less than 1.5 µmol/L/h (and the CV could be as high as 69%). FDA estimates that the LoQ based on an imprecision goal of 20% CV (which is the typical imprecision goal for the LoQ of quantitative assays) is approximately 4.7 µmol/L/h and above the claimed LoQ of the assay. Because the LoD of the assay could be as high as 5.36

µmol/L/h, the true LoQ of this assay is somewhere between 4.7 and 5.36 µmol/L/h (note that one of the high risk cutoffs for this assay is 4.5 µmol/L/h).

Outliers: Baebies identified several statistical outliers in their analytical studies in support of the GAA assay (and all assays). MSPHL also identified "visual outliers" during the clinical study when clinical samples were tested in multiple replicates. In the analytical studies (e.g., linearity, detection limits) these outliers impacted the performance of the GAA assay.

Stability of the GAA samples (transport stability): Baebies performed a study showing the impact of shipping the DBS for up to 5 days at the following ambient conditions: 10ºC and 20% relative humidity (RH), 10ºC and 80% RH, 45ºC and 20% RH, 45ºC and 80% RH and 25ºC and 50% RH.

The impact on the test result is summarized in the table below:

Average GAA enzymatic activity for DBS samples before and after simulated transport

Condition	Day	Sample 3 Concentration µmol/L/hr		Sample 5 Concentration µmol/L/hr		Sample 10 Concentration µmol/L/hr		Sample 15 Concentration µmol/L/hr	
		before	after	before	after	before	after	before	after
10C, 20%RH	1	5.9	5.7	6.7	6.6	13.9	14.2	26.5	24.4
	3	5.7	5.7	6.7	6.6	13.3	13.5	27.0	28.6
	5	6.5	6.2	7.2	6.8	15.0	15.5	28.5	28.7
10C, 80%RH	1	6.6	6.6	7.2	7.2	14.7	14.8	27.2	26.7
	3	5.9	6.0	6.5	6.6	13.4	14.1	25.4	26.2
	5	6.2	5.7	7.1	6.8	14.2	14.0	29.0	29.5
25C, 50%RH	1	7.5	6.5	7.3	7.0	16.3	16.6	27.5	24.8
	3	6.3	6.6	7.5	7.4	15.0	14.7	26.8	25.0
	5	6.4	6.4	7.4	7.5	14.7	15.2	27.6	24.3
45C, 20%RH	1	6.6	5.5	6.9	6.2	14.3	13.0	26.3	21.6
	3	6.1	5.5	7.3	6.1	14.1	11.8	27.9	19.5
	5	6.5	6.0	7.7	6.5	14.9	12.3	26.4	18.0
45C, 80%RH	1	6.5	5.2	7.2	5.7	16.1	11.3	28.2	18.6
	3	6.1	3.9	7.1	4.3	13.7	6.3	26.3	7.9
	5	6.2	3.5	7.0	3.3	14.3	4.4	28.8	4.6

The samples above presented in red font are outside the acceptance criteria for acceptable stability defined by Baebies (i.e., recovery values <85% and >115% for enzyme concentrations greater than ≈6 µmol/L/hr were considered to be significantly impacted). The highlighted fields show samples with targeted concentrations set to represent normal and high normal samples that upon transport at 45ºC and 80% relative humidity fall either below the borderline cutoff or some high risk cutoffs (borderline cutoff set to 10 µmol/L/hr and high risk cutoff ranged from 4.5 to 7.2 µmol/L/hr for this assay). These shipping conditions might be experienced in certain regions of the United States such as in the southern states especially if shipped at ambient conditions.

This study demonstrates that the enzyme activity will decrease when exposed to high temperature and/or humidity and should be taken into account in certain geographic areas/seasons. This degradation could lead to false positive results. FDA is asking for panel input on whether the panel is aware of any measures that Baebies

can recommend in their Instructions for Use to mitigate the impact on the enzyme activity as result of standard shipping conditions.

Analytical Performance of the Assay at the Cutoffs

In summary:

- The data in support of the LoD and LoQ of this assay is difficult to interpret and it appears that the highlighted cutoffs could fall below the LoD of the assay.

GAA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
Pivotal 1	5/16/2013	7.0	7.0	4.5	10.0	10.0	10.0
Pivotal 2	11/12/2013	7.2	7.2	4.5	10.0	10.0	10.0
Pivotal 3	6/23/2014	7.2	4.5	4.5	10.0	10.0	10.0

- When samples from babies confirmed positive for Pompe were retested by MSPHL as part of their quality assurance, between 8.6 and 11.5% of the test results fell above the high risk cutoffs upon retesting.

- There were more outliers than is typical for assays of this type (statistical outliers during the analytical studies and visual outliers during the clinical study).

- Transport at certain ambient conditions resulted in decreased enzyme activity. We are aware that laboratories have reported seasonal variation with other analyte levels.

To conclude, for this clinical use, newborn screening test manufacturers typically provide clinical data in the package insert that is analyzed with cutoffs set so that no known positive babies are incorrectly categorized (i.e., presumed normal instead of high risk for the screened condition and vice versa) and with cutoffs that are also analytically reliable.

To summarize FDA's questions for the panel, we are seeking input on:

- Whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis) or the false positive rate estimate based on the test alone, or another specified analysis?

- Whether the panel has a recommendation on how to estimate the false negative rate of this device?

- Whether the analysis of the clinical study should use the cutoffs used to test each baby during the study or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff)? This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

- Whether performance characteristics for age-specific cutoffs, when used, should be provided in the Instructions for use?

- Whether there are any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity as result of standard shipping conditions, including high temperature and humidity?

- Whether the analytical performance of the assays at the cutoffs (e.g., precision, detection limits, outliers, performance of confirmed positive samples upon retesting) is sufficient for safe and effective use (i.e., does the panel consider that Baebies has demonstrated adequate analytical validity of the assays)?

D. Screening Results of the GBA (β-glucocerebrosidase) Assay to Screen for Gaucher During the Pivotal Study Phase

Next we present the results of the GBA enzyme activity assay intended to screen babies for Gaucher. For this study, two cutoffs – a high risk cutoff and a borderline cutoff – were also used. Similar to the GAA assay, the high-risk cutoffs were different depending on the age of the baby at the time the screen was collected and were developed to account for the change in enzyme activity (see table titled "Reference intervals for GBA" below) as a function of age at the time of sample collection (for a detailed discussion on the changes, please refer to Appendix A provided by Baebies "Age At Collection Related Changes In Activity"). However, these age-specific cutoffs were added during the pilot and pivotal phases of the study. MSPHL added a different cutoff for newborns ≥ 14 days of age in May 16, 2013 and a different cutoff for newborns 7-13 days of age on June 23, 2014 as the laboratory noted these changes in the enzyme levels due to age of the newborn when the specimen was collected.

Reference intervals for GBA (calculated from the presumed normal babies)

		Quantile (µmol/L/hr)		Population Activities (µmol/L/hr)			
Age	n	0.10%	99.90%	Q1	Mean	Median	Q3
1-6 days	151,960	6.10	68.00	15.81	20.92	19.70	24.54
7-13 days	10,620	4.61	80.95	12.97	17.69	16.15	20.39
14+ days	12,880	4.07	79.26	11.14	15.65	14.20	18.00

The following cutoffs were used during clinical study:

GBA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	4.5	4.5	4.5	7.0	7.0	7.0
2	3/9/2013	7.0	7.0	7.0	8.0	8.0	8.0
3	5/16/2013	7.0	7.0	5.0	8.0	8.0	8.0
4*	6/5/2013	5.5	5.5	5.0	8.0	8.0	8.0
5	8/28/2013	5.5	5.5	5.0	7.0	7.0	7.0
6	6/23/2014	5.5	4.0	4.0	7.0	7.0	7.0

* This change in cutoff occurred during the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

Baebies provided the following summary table of the screening results for the GBA assay during the pivotal phase of the study. This table summarized the results using the cutoffs that MSPHL was using at the time the samples were tested (i.e., though Missouri changed the cutoffs during the study, this data is analyzed using the cutoffs used at the time the newborn screen of each baby was performed).

		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,688	9,149	2,148	351	104,336
	At least one screen with first result below borderline	436	191	105	21	753
Average of all Tests	All screens w/ avg. above high risk	411	168	94	17	690
	At least one screen w/ avg. below high risk	25	23	11	4	63
	Referred	6	2	-	-	8
	Not Referred	19	21	11	4	55
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Gaucher - True Positive	2				2
	Gaucher - Normal, False Positive	1	2			3
	Gaucher - Carrier, False Positive	2				2
	Gaucher - Refused	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		11	6	3	20
	Later Sample From Newborn Above Cutoff		9	4	1	14
	Other Assay Below Borderline	10	1			11
	Outliers Excluded	6				6
	Transfused	2		1		3
	Other Assay Low-Normal	1				1
Performance Summary	Total Presumed Normal	93,118	9,338	2,253	372	105,081
	Total Presumed Affected	6	2	-	-	8
	True Positives	2	0	0	0	2
	False Positives	3	2	-	-	5
	False Positive Rate (FPR)	0.003%	0.021%	0.000%	0.000%	0.005%

For the GBA assay the table above shows that overall 105,089 newborns were included in the analysis during the pivotal study phase (as defined by Baebies post-hoc; see description on page 12 of the post-study division into pilot and pivotal phases). Of these:

- 104,336 neonates' initial test results were considered low risk for Gaucher and they were presumed normal.

- 753 babies' initial test results indicated an undetermined risk for Gaucher and they were subjected to additional testing.

Of these 753 newborns, 690 babies' repeat test results indicated a low risk for Gaucher and they were presumed normal.

The remaining 63 neonates had results that indicated that they were at high risk for Gaucher. These newborns were subjected to the risk assessment described above in table 1. Based on the results of the risk assessment, of the 63 babies with high risk test results:

- 8 babies were still considered high risk for Gaucher and were referred to the metabolic centers for confirmatory diagnostic testing, and
- 55 babies were not referred because after the risk assessment they were not considered to be at high risk for Gaucher.

Of the 8 babies that were referred for diagnostic follow-up:

- 2 were diagnosed with Gaucher
- 2 were determined to be carriers
- 3 were determined to be normal, and
- 1 family refused follow-up.

Meanwhile for the 55 babies that were not referred the following factors were considered to reduce their risk of Gaucher:

- 20 babies had prior test results (from prior screens) that indicated that they were at low risk for Gaucher. 14 babies had subsequent tests (from a subsequent screen) that indicated that the baby was at low risk for Gaucher. As discussed above in the risk assessment section and in the IDUA section, during the risk assessment MSPHL evaluated all test results from all tests and retests from all screens and re-screens available for the baby to make their final risk determination. Again, it is unclear if this same number of convenience-sample test results would be available to other laboratories/programs implementing this test system.
- 6 babies were not referred because MSPHL identified visual outliers among the test results. Once the visual outliers were excluded from the analysis, the average of the remaining test results indicated that the babies were at low risk for Gaucher and they were presumed normal.
- 11 babies were not referred becasue the other LSD assay results indicated high risk for those screened conditions.
- 1 baby had 1 LSD assay result that was low-normal. This was considered very unlikely, so the samples were presumed to be of low quality. Per MSPHL's risk analysis, this reduced the risk for Gaucher and these babies were presumed normal.
- 3 babies were not referred because they were transfused.

A summary of the screening performance of the GBA assay for all babies screened during the retrospectively-defined pivotal phase of the study is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Gaucher (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GBA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,099	9,317	2,242	368	105,026
Test screen positive	25	23	11	4	63
Referred	6	2	0	0	8
Not referred	19	21	11	4	55
True positives	2	0	0	0	2

Estimated False Positive Rates

Baebies estimated the false positive rate of MSPHL's screening procedure as 0.005% using the following formula:

> The number of false positives (5) from the babies referred for diagnostic follow up (not including babies lost to follow-up)
>
> divided by
>
> The total number of babies screened (105,089) minus any confirmed positive babies (2) and minus any babies lost to follow up (1)
>
> That is 5/105,086=0.005%

To calculate the false positive rate of the assay on the Seeker System, FDA used a different approach and the following formula:

> The number of babies with test results indicating high risk of Gaucher (i.e., below the "high-risk" cutoff) that did not have Gaucher (63) minus the number of babies lost to follow-up (1)
>
> divided by
>
> The total number of newborn screened (105,089) minus the number of confirmed affected babies (2) minus any babies lost to follow up (1)
>
> That is 62/105,086=0.059%

As discussed above for the other assays, the difference between the false positive rate calculated for the MSPHL's screening procedure (0.005%) and the false positive rate of the assay calculated by FDA (0.059%) is that the false positive rate of the assay includes the babies with test results that indicated a high risk for Gaucher that were not referred because the risk analysis lowered the risk for the baby. The reason for including these babies in the false positive rate for the device is that the risk analysis is a component of the laboratory's practice based on the expertise of the laboratory and is independent of the device. Manufacturers of newborn screening tests include false positive rate estimates for their device in their Instructions for Use so laboratories who install and use their system will know what to expect regarding this rate when implementing their testing. FDA would like panel input on whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or other formats of information to assist laboratories in implementing this test.

Estimated False Negative Rate

Based on the newborn screening follow-up program, Baebies reports no known false negatives for the GBA assay since there have been no babies that were screened during the study period referred to the metabolic centers and diagnosed with Gaucher.

To provide additional information on the false negative rate, Baebies performed a retrospective analysis of the initial screening and repeat testing of confirmed positive Gaucher samples as well as additional testing of the samples for quality assurance purposes. In this analysis, 15 tests from several confirmed positive Gaucher samples tested during the course of the study were evaluated. The results of this analysis are provided below:

Table 5 - Results Above Cutoff Based on Cutoff Used – GBA – all samples > 24h

Cutoff	# of Tests of Affected Samples	# of Tests above cutoff	% of tests above cutoff
High Risk (at time of test)	15	3	20.0%
Borderline (at time of test)	15	0	0.0%
High Risk (final)	15	3	20.0%
Borderline (final)	15	0	0.0%

The result of each test was evaluated based on the cutoffs that were in use when the sample was tested (and described above as "at the time of test") and the cutoffs that were in use at the end of the study (and described above as "final"). No test results were above the borderline cutoff (presumed normal), however 20% fell above both the high risk cutoff (in the borderline region between high risk and borderline cutoff) used at the time the

sample was tested and the final high risk cutoff used at the end of the study. There was no root cause analysis given for the analytical false negatives detected.

FDA would like panel input on whether this type of analysis is adequate to conclude that the GBA assay has no false negatives.

Information on Confirmed Positive Babies

During the retrospectively-defined pivotal clinical study, 2 newborns were diagnosed with Gaucher. All 2 confirmed positive babies were identified and referred following the first valid screen.

Modifications to the Cutoffs During the Clinical Study

The GBA cutoffs were modified several times by MSPHL during the pilot phase of the study and once during the pivotal phase.

GBA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	4.5	4.5	4.5	7.0	7.0	7.0
2	3/9/2013	7.0	7.0	7.0	8.0	8.0	8.0
3	5/16/2013	7.0	7.0	5.0	8.0	8.0	8.0
4*	6/5/2013	5.5	5.5	5.0	8.0	8.0	8.0
5	8/28/2013	5.5	5.5	5.0	7.0	7.0	7.0
6	6/23/2014	5.5	4.0	4.0	7.0	7.0	7.0

* This change in cutoff occurred before the end of the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

The timing, reason and a description of the change provided by MSPHL to the GBA high risk cutoffs (during the entire clinical study (i.e., pilot and pivotal) are summarized in the table below:

Date of cutoff change	Cutoff in use	Summary of information reviewed	Cutoff decision
3/9/2013	4.5µmol/L/hr	Referral rate – 0% There were no referrals so risk of false negative is higher.	Increase from 4.5 to 7.0 µmol/L/hr
5/16/2013	7.0 µmol/L/hr	The distribution of the 14+ day babies was evaluated and a different cutoff was introduced. The high risk 14+ cutoff was set to the same percentile as the cutoff percentile (i.e., 0.12^{th}) for the 1-6 day babies	14+ day high risk cutoff introduced and reduced from 7.0 to 5.0 µmol/L/hr
06/05/2013	7.0µmol/L/hr	Referral rate – 0.049% False positive rate – 0.049% No cases diagnosed.	Decrease cutoff used for all babies less than 14 days old from 7.0 to 5.5 µmol/L/hr

Date of cutoff change	Cutoff in use	Summary of information reviewed	Cutoff decision
6/23/2013		The distribution of the 7-13 day babies was evaluated and a different cutoff was introduced. The high risk 7-13 day cutoff was set to the same percentile as the cutoff percentile (i.e., 0.05th) for the 1-6 day babies. 14+ day cutoff was also set to the same percentile as the cutoff percentile (i.e., 0.05th) for the 1-6 day babies	7-13 day high risk cutoff introduced and reduced from 5.5 to 4.0 µmol/L/hr 14+ cutoff reduced from 5.0 to 4.0 µmol/L/hr

To explore the impact of the changes to the cutoffs during the pivotal study, Baebies provided a table summarizing the screening performance of the device for the entire pivotal phase of the study applying the final cutoffs used by MSPHL. The screening results are summarized in the table below:

GBA - Pivotal Phase						
		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,688	9,149	2,148	351	104,336
	At least one screen with first result below borderline	436	191	105	21	753
Average of all Tests	All screens w/ avg. above high risk	413	173	97	18	701
	At least one screen w/ avg. below high risk	23	18	8	3	52
	Referred	6	2	-	-	8
	Not Referred	17	16	8	3	44
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Gaucher - True Positive	2				2
	Gaucher - Normal, False Positive	1	2			3
	Gaucher - Carrier, False Positive	2				2
	Gaucher - Refused	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		6	4	2	12
	Later Sample From Newborn Above Cutoff		9	4	1	14
	Other Assay Below Borderline	9	1			10
	Outliers Excluded	6				6
	Transfused	1				1
	Other Assay Low-Normal	1				1
Performance Summary	Total Presumed Normal	93,118	9,338	2,253	372	105,081
	Total Presumed Affected	6	2	-	-	8
	True Positives	2	0	0	0	2
	False Positives	3	2	-	-	5
	False Positive Rate (FPR)	0.003%	0.021%	0.000%	0.000%	0.005%

The impact of the changes to the cutoffs is that 11 babies that were considered high risk for Gaucher (based on the cutoffs in use by MSPHL at the time these babies were screened) would no longer be considered high risk (i.e., during routine screening, 63 babies were considered high risk for Gaucher based on the test result and using the final cutoffs, 52 babies would be considered high risk for Gaucher). The false positive rate for MSPHL's screening procedure is the same (0.005%) and the false positive rate of the test (i.e., FDA's analysis) also remains consistent at 0.047% using the final cutoffs compared to 0.059% based on the cutoffs used at testing. All confirmed positive Gaucher babies (including those diagnosed during the pilot phase of the study) would still be considered high risk for Gaucher if the final cutoffs (and all cutoffs used during the study) are applied to the original test results.

FDA would like the panel's input on whether the analysis of the clinical study should use the cutoffs used when testing each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final

cutoff – cutoff period 6). This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

A summary of the screening performance of the GBA assay for all babies screened during the retrospectively-defined pivotal phase of the study is provided below, using the final cutoffs. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Gaucher (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GBA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,101	9,322	2,245	369	105,037
Test screen positive	23	18	8	3	52
Referred	6	2	0	0	8
Not referred	17	16	8	3	44
True positives	2	0	0	0	2

Analytical Performance of the GBA Assay

During our review of the analytical performance data provided in support of the GBA assay we noted the following:

Precision: Baebies performed a study to estimate the imprecision of the GBA assay using 4 analyzers and 3 reagent lots and performed testing during 21 non-consecutive days, with 2 runs per day and 2 dried blood spot (DBS) punches of each specimen per run. Baebies tested a total of 336 replicate results for each specimen. In this study 11 samples were identified as invalid. The results of the study are summarized in the table below. In the following table the repeatability (or within-run precision) estimate includes the instrument run as the component (or condition) of variability and the reproducibility estimate includes the instrument run, reagent lot, instrument and day as the components (or conditions) of variability. As a reminder, the final cutoffs during the pivotal study were 5.5 and 4.0 µmol/L/h (high risk cutoffs) and 7.0 µmol/L/h (borderline cutoff).

Mean µmol/L/h	N	Repeatability (%CV)	Between Lot (%CV)	Between Instrument (%CV)	Between Day (%CV)	Reproducibility (%CV)
2.84	331	34.9%	12.7%	1.8%	2.5%	38.0%
3.47	334	13.5%	10.7%	2.3%	0.0%	18.4%
5.07	335	11.0%	11.2%	0%	0.0%	16.6%
8.55	334	11.6%	10.4%	1.4%	0.0%	15.8%
15.00	335	11.3%	11.4%	2.1%	1.2%	15.7%

The coefficient of variability (CV) was calculated for each imprecision estimate. The CV is calculated by dividing the standard deviation by the mean of the sample and describes the extent of variability in relation to the mean of the sample. The highlighted sections in the table above denote the imprecision of the GBA assay near the cutoffs used in the pivotal study

Detection Limits: Baebies evaluated the detection limits of this assay following a recognized guideline[10] using 3 lots of their reagent and estimated the following detection limits:

- The limit of the blank (LoB) was defined as the highest analyte concentrations expected to be found when replicates of a sample containing no analyte are tested with 95% confidence. This is often a way of determining what concentration(s) the assay cannot distinguish from "noise." Baebies estimated that the LoB of the GBA assay was 0.72 µmol/L/h.

- The limit of detection (LoD) was defined as the lowest analyte concentration likely to be reliably distinguished from a blank sample with 95% confidence. The LoD of the GBA assay was determined to be 1.07 µmol/L/h.

- The limit of quantification was defined as the lowest concentration where the total imprecision was SD ≤ 1.5 µmol/L/h or 20% CV whichever is greater. Baebies estimated that the LoQ for the GBA assay was 1.85 µmol/L/h which was the concentration where the imprecision was SD less than 1.5 µmol/L/h (and the %CV could be as high as 81%). FDA estimates that the LoQ based on an imprecision goal of 20% CV (which is the typical imprecision goal for the LoQ of quantitative assays) is approximately 3 µmol/L/h.

Outliers: Baebies identified several statistical outliers in their analytical studies in support of the GBA assay (and all assays). As discussed above for the other assays, MSPHL also identified "visual outliers" during the clinical study when clinical samples needed to be tested in multiple replicates. In the analytical studies (e.g., detection limits, linearity) these outliers affected the performance of the assay.

Stability of the GBA samples (transport stability): Baebies performed a study showing the impact of shipping the DBS for up to 5 days at the following ambient conditions: 10ºC and 20% relative humidity (RH), 10ºC and 80% RH, 45ºC and 20% RH, 45ºC and 80% RH and 25ºC and 50% RH.

The impact on the test result is summarized in the table below:

Average GBA enzymatic activity for DBS samples before and after simulated transport

Condition	Day	Sample 3 Concentration µmol/L/hr		Sample 5 Concentration µmol/L/hr		Sample 10 Concentration µmol/L/hr		Sample 15 Concentration µmol/L/hr	
		before	after	before	after	before	after	before	after
10C, 20%RH	1	3.5	3.6	3.8	3.8	6.8	6.6	11.7	10.7
	3	3.7	3.6	3.6	3.6	6.3	6.9	11.7	11.6
	5	3.9	3.7	4.2	4.2	6.8	7.0	12.2	12.2
10C, 80%RH	1	3.9	4.0	4.2	4.1	7.0	7.6	12.2	11.4
	3	4.0	4.0	4.1	4.0	6.6	6.6	11.1	11.2
	5	3.7	3.1	4.0	3.8	6.7	6.3	11.6	10.7
25C, 50%RH	1	4.2	3.7	4.0	3.9	7.2	7.2	11.7	10.0
	3	3.9	3.9	4.1	4.2	7.0	6.5	11.6	10.1
	5	4.0	3.8	4.3	4.2	7.0	6.5	11.2	9.9
45C, 20%RH	1	3.5	3.4	3.8	3.6	6.8	5.6	11.5	7.7
	3	3.7	3.9	4.3	3.8	6.8	5.5	12.3	7.0
	5	3.6	4.3	4.3	4.2	6.4	6.3	10.8	5.9

Condition	Day	Sample 3 Concentration µmol/L/hr		Sample 5 Concentration µmol/L/hr		Sample 10 Concentration µmol/L/hr		Sample 15 Concentration µmol/L/hr	
		before	after	before	after	before	after	before	after
45C, 80%RH	1	3.9	3.8	4.2	3.7	7.7	5.3	12.2	6.8
	3	4.2	3.5	4.7	3.5	6.5	4.3	11.2	4.8
	5	3.8	3.6	4.1	3.2	7.1	3.5	11.7	3.6

The samples above presented in red font are outside the acceptance criteria for acceptable stability defined by Baebies (i.e., recovery values <85% and >115% for enzyme concentrations greater than ≈6 µmol/L/hr and mean difference in enzymatic activity outside +/-1 µmol/L/h for enzyme concentrations below ≈6 µmol/L/hr were considered to be significantly impacted). The highlighted fields show samples with targeted concentrations set to represent normal samples that upon transport at 45ºC and 20% and 80% relative humidity fall either below the borderline cutoff or some high risk cutoffs (borderline cutoff set to 7 µmol/L/hr and high risk cutoff ranged from 4.0 to 5.5 µmol/L/hr for this assay). These shipping conditions might be experienced in certain regions of the United States such as in the southern states especially if shipped at ambient conditions.

This study demonstrates that the enzyme activity will decrease when exposed to high temperature and/or humidity, and should be taken into account in certain geographic areas/seasons. This degradation could lead to false positive results. FDA is asking for panel input on whether the panel is aware of any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity as result of standard shipping conditions.

Analytical Performance of the Assay at the Cutoffs

In summary:

- When samples from babies confirmed positive for Gaucher were retested by MSPHL as part of their quality assurance, 20% of the test results fell above the high risk cutoffs upon retesting.

- There were more outliers than is typical for assays of this type (statistical outliers during the analytical studies and visual outliers during the clinical study).

- Transport at certain ambient conditions resulted in decreased enzyme activity. We are aware that laboratories have reported seasonal variation with other analyte levels.

To conclude, for this clinical use, newborn screening test manufacturers typically provide clinical data in the package insert that is analyzed with cutoffs set so that no known positive babies are incorrectly categorized (i.e., presumed normal instead of high risk for the screened condition and vice versa) and with cutoffs that are also analytically reliable.

To summarize FDA's questions for the panel, we are seeking input on:

- Whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or another specified analysis?

- Whether the panel has a recommendation on how to estimate the false negative rate of this device?

- Whether the analysis of the clinical study should use the cutoffs used to test each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff)? This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

- Whether performance characteristics for age-specific cutoffs, when used, should be provided in the Instructions for use?

- Whether there are any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity standard shipping conditions, including high temperature and humidity?

- Whether the analytical performance of the assays at the cutoffs (e.g., precision, detection limits, outliers, performance of confirmed positive samples upon retesting) is sufficient for safe and effective use (i.e., does the panel consider that Baebies has demonstrated adequate analytical validity of the assays)?

E. Screening Results for α-D-galactosidase A (GLA) Assay to Screen for Fabry During the Pivotal Study Phase

Lastly we present the results of the GLA enzyme activity assay intended to screen babies for Fabry. For this study, two cutoffs – a high risk cutoff and a borderline cutoff – were used. Similar to the GAA and the GBA assays, the high-risk cutoffs were different depending on the age of the newborn at the time the screen was collected to account for the change in enzyme activity (see table titled "Reference intervals for GLA" below) as a function of the age of the baby at the time of sample collection (refer to Baebies' Appendix A for additional information). However, these age-specific cutoffs were added during the pilot and pivotal phases of the study. MSPHL added a different cutoff for newborns ≥ 14 days of age in May 16, 2013 and a different cutoff for newborns 7-13 days of age on April 23, 2014 as the laboratory noted these changes in the enzyme levels due to age of the newborn when the specimen was collected. In contrast to the GAA and GBA assays, the borderline cutoffs also differed depending on the age of the newborn at the time the screen was taken; this change was made on April 23, 2014.

Reference intervals for GLA (calculated from the presumed normal babies)

		Quantile (µmol/L/hr)		Population Activities (µmol/L/hr)			
Age	n	0.10%	99.90%	Q1	Mean	Median	Q3
1-6 days	151,960	6.70	180.15	18.69	29.86	25.32	35.31
7-13 days	10,620	4.45	129.75	14.01	22.66	19.03	26.65
14+ days	12,880	3.41	120.75	10.45	16.82	14.01	19.41

The following cutoffs were sued during the clinical study:

GLA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	5.5	5.5	5.5	7.0	7.0	7.0
2	3/9/2013	6.2	6.2	6.2	7.5	7.5	7.5
3	5/16/2013	6.2	6.2	3.7	7.5	7.5	7.5
4*	6/5/2013	7.0	7.0	3.7	8.0	8.0	8.0
5	12/17/2013	8.0	8.0	3.7	10.0	10.0	10.0
6	4/23/2014	8.0	5.0	3.7	10.0	5.0	5.0

GLA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
7	6/4/2014	7.0	5.0	3.7	9.0	5.0	5.0
8	6/23/2014	7.0	5.0	3.0	9.0	5.0	5.0

* This change in cutoff occurred in the middle of the retrospectively-defined pilot phase; this cutoff was used at the beginning of the retrospectively-defined pivotal phase. As stated above the pivotal phase was redefined as starting on August 27, 2013.

Baebies provided the following summary table of the screening results for the GLA assay during the pivotal phase of the study. This table summarized the results using the cutoffs that MSPHL was using at the time the samples were tested (i.e., though Missouri changed the cutoffs during the study, this data is analyzed using the cutoffs used at the time the newborn screen of each baby was performed).

		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,405	8,913	2,032	320	103,670
	At least one screen with first result below borderline	719	427	221	52	1,419
Average of all Tests	All screens w/ avg. above high risk	617	360	193	49	1,219
	At least one screen w/ avg. below high risk	102	67	28	3	200
	Referred	50	9	1	-	60
	Not Referred	52	58	27	3	140
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Fabry - True Positive	26	4			30
	Fabry - Normal, False Positive	20	5	1		26
	Fabry - Refused	3				3
	Fabry - Moved	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		43	24	2	69
	Other Assay Below Borderline	20	3	1		24
	Later Sample From Newborn Above Cutoff	1	12	2		15
	Outliers Excluded	14				14
	Transfused	5			1	6
	Different Cutoff Applied	3				3
	Spot Variability	3				3
	Retrospectively Referred	2				2
	Age Related Enzyme Decrease	1				1
	Contaminated Sample	1				1
	Multiple Reasons	1				1
	Other Assay Low-Normal	1				1
Performance Summary	Total Presumed Normal	93,074	9,331	2,252	372	105,029
	Total Presumed Affected	50	9	1	-	60
	True Positives	26	4	0	0	30
	False Positives	20	5	1	-	26
	False Positive Rate (FPR)	0.021%	0.054%	0.044%	0.000%	0.025%

For the GLA assay the table above shows that overall 105,089 babies were included in the analysis during the pivotal study phase (as defined by Baebies post-hoc; see description on page 12 of the post-study division into pilot and pivotal phases). Of these:

- 103,670 babies' initial test results were considered low risk for Fabry and they were presumed normal.
- 1,419 babies' initial test results indicated an undetermined risk for Fabry and they were subjected to additional testing.

Of these 1,419 babies, 1,219 babies' repeat test results indicated a low risk for Fabry and they were presumed normal.

The remaining 200 babies had results that indicated that they were at high risk for Fabry. These babies were subjected to the risk assessment described above in table 1. Based on the results of the risk assessment:

- 60 babies were still considered to be at high risk for Fabry and were referred to the metabolic centers for confirmatory diagnostic testing, and
- 140 babies were not referred because after the risk assessment they were not considered to be at high risk for Fabry.

Of the 60 babies that were referred for diagnostic follow-up:

- 30 were diagnosed with Fabry
- 26 were determined to be normal, and
- 4 were lost to follow-up.

Meanwhile for the 140 babies that were not referred the following factors were considered to reduce their risk of Fabry:

- 69 babies had prior test results (from prior screens) that indicated that they were at low risk for Fabry. 15 babies had subsequent tests (from a subsequent screen) that indicated that the baby was at low risk for Fabry. As discussed above in the risk assessment section and in the IDUA section, during the risk assessment MSPHL evaluated all test results from all tests and retests from all screens and re-screens available for the baby to make their final risk determination. However, it is unclear if this same number of convenience-sample test results would be available to other laboratories/programs implementing this test system.
- 14 babies were not referred because MSPHL identified visual outliers among the test results. Once the visual outliers were excluded from the analysis, the average of the remaining test results indicated that the babies were at low risk for Fabry and they were presumed normal.
- 24 babies not referred because the other LSD assay results were below the borderline cutoffs for other screened conditions.
- 1 baby had 1 LSD assay result that was low-normal. This was considered very unlikely, so the samples were presumed to be of low quality. Per MSPHL's risk analysis, this reduced the risk for Fabry and these babies were presumed normal.
- 6 babies were not referred because they were transfused
- 11 babies were not referred for the following reasons:
    - different cutoff applied (3)
    - spot variability (3)
    - retrospectively referred (2)
    - age related enzyme decrease (1)
    - contaminated sample (1)
    - "multiple reasons" (1)

A summary of the screening performance of the GLA assay for all babies screened during the retrospectively-defined pivotal phase of the study is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Fabry (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GLA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,022	9,273	2,225	369	104,889
Test screen positive	102	67	28	3	200
Referred	50	9	1	0	60
Not referred	52	58	27	3	140
True positives	26	4	0	0	30

Estimated False Positive Rates

Baebies estimated the false positive rate of MSPHL's screening procedure as 0.025% using the following formula:

> The number of false positives (26) from the babies referred for diagnostic follow up (not including babies lost to follow-up)
>
> divided by
>
> The total number of babies screened (105,089) minus any confirmed positive babies (30) and minus any babies lost to follow up (4)
>
> That is 26/105,055=0.025%

To calculate the false positive rate of the assay on the Seeker System, FDA used a different approach and the following formula:

> The number of babies with test results indicating high risk of Fabry (i.e., below the "high-risk" cutoff) that did not have Fabry (170) minus the number of babies lost to follow-up (4)
>
> divided by
>
> The total number of newborn screened (105,089) minus the number of confirmed affected babies (30) minus any babies lost to follow up (4)
>
> That is 166/105,055=0.16%

As discussed above for the other assays, the difference between the false positive rate calculated for MSPHL's screening procedure (0.025%) and the false positive rate of the assay calculated by FDA (0.16%) is that the false positive rate of the assay includes the babies with test results that indicated a high risk for Fabry that were not referred because the risk analysis lowered the risk for the baby. The reason for including these babies in the false positive rate for the device is that the risk analysis is a component of the laboratory's practice based on the expertise of the laboratory and is independent of the device.

Manufacturers of newborn screening tests include false positive rate estimates for their device in their Instructions for Use so laboratories who install and use their system will know what to expect regarding this rate when implementing their testing.

FDA would like panel input on whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or other formats of information to assist laboratories in implementing this test.

Estimated False Negative Rate

Based on the newborn screening follow-up program, Baebies reports no known false negatives for the GLA assay since there have been no babies that were screened during the study period referred to the metabolic centers and diagnosed with Fabry.

To provide additional information on the false negative rate, Baebies performed a retrospective analysis of the initial screening and repeat testing of confirmed positive Fabry samples as well as additional testing of the samples for quality assurance purposes. In this analysis, 285 tests from several confirmed positive Fabry samples tested during the course of the study were evaluated. The results of this analysis are provided below:

Table 6 - Results Above Cutoff Based on Cutoff Used – GLA – all samples > 24h

Cutoff	# of Tests of Affected Samples	# of Tests above cutoff	% of tests above cutoff
High Risk (at time of test)	285	61	21.4%
Borderline (at time of test)	285	13	4.6%
High Risk (final)	285	64	22.5%
Borderline (final)	285	18	6.3%

Baebies evaluated the result of each test based on the cutoffs that were in use when the sample was tested (and described above as "at the time of test") and the cutoffs that were in use at the end of the study (and described above as "final"). Between 4.6 and 6.3% of the confirmed positive Fabry samples were above the borderline cutoff (i.e., presumed normal) upon retesting and between 21.4 and 22.5% of the results previously categorized as screen positive (i.e., below the high risk cutoff) were above the high risk cutoff upon retesting. There was no root cause analysis given for the analytical false negatives detected.

FDA would like panel input on whether this type of analysis is adequate to conclude that there are no false negatives for the GLA assay.

Information on Confirmed Positive Babies

During the pivotal clinical study, 30 babies were diagnosed with Fabry.
Four (4) confirmed positive Fabry babies had multiple screens. These babies were all identified and referred following the first valid screen.

Modifications to the Cutoffs During the Clinical Study

The GLA cutoffs were modified several times by MSPHL during the pilot phase of the study and once during the pivotal phase. The following cutoffs were used during clinical study:

GBA cutoffs		High Risk			Borderline Risk		
Cutoff period	Effective Date	0-6 days	7-13 days	14+ days	0-6 days	7-13 days	14+ days
1	1/15/2013	5.5	5.5	5.5	7.0	7.0	7.0
2	3/9/2013	6.2	6.2	6.2	7.5	7.5	7.5
3	5/16/2013	6.2	6.2	3.7	7.5	7.5	7.5
4*	6/5/2013	7.0	7.0	3.7	8.0	8.0	8.0
5	12/17/2013	8.0	8.0	3.7	10.0	10.0	10.0
6	4/23/2014	8.0	5.0	3.7	10.0	5.0	5.0
7	6/4/2014	7.0	5.0	3.7	9.0	5.0	5.0
8	6/23/2014	7.0	5.0	3.0	9.0	5.0	5.0

* This change in cutoff occurred before the end of the retrospectively-defined pilot phase and was used at the start of the retrospectively-defined pivotal phase; as stated above the pivotal phase was redefined as starting on August 27, 2013.

The timing, reason and a description of the changes provided by MSPHL for the GLA high risk cutoffs (during the entire clinical study (i.e., pilot and pivotal) are summarized in the table below:

Date of change to Cutoff	Cutoff in use	Summary of information reviewed	Cutoff decision
03/08/2013	5.5µmol/L/hr	Lab's Referral rate – 0.015% *Lab's False positive rate – 0% Low referrals and no false positives. Risk of false negatives is higher.	Increase from 5.5 to 6.2 µmol/L/hr
5/16/2013	6.2 µmol/L/hr	The distribution of the 14+ day babies was evaluated and a different cutoff was introduced. The high risk 14+ cutoff was set to the same percentile as the cutoff percentile (i.e., 0.15^{th}) for the 1-6 day babies	14+ day high risk cutoff introduced and reduced from 6.2 to 3.7 µmol/L/hr
06/05/2013	6.2µmol/L/hr	Lab's Referral rate – 0.088% *Lab's False positive rate – 0.034% PPV – 61% False positive rate was considered to be still low given the high PPV of this test. Risk of false negatives is higher.	Increase 1-6 day and 6-13 day cutoffs from 6.2 to 7.0 µmol/L/hr

Date of change to Cutoff	Cutoff in use	Summary of information reviewed	Cutoff decision
12/17/2013	7.0µmol/L/hr	GLA percentile values (1st and median) were trending higher resulting in lower retest and referral rate. Though the shift was originally attributed to a reagent lot change, review of the percentiles indicates that the trend started before the reagent lot change suggesting that weather change (cooler weather) was the primary contributor. The effect of seasonal changes in weather on lysosomal enzyme activity was not well understood at that time.	Increase Increase 1-6 day and 6-13 day cutoffs from 7.0 to 8.0 µmol/L/hr
4/23/2014	8.0 µmol/L/hr	The distribution of the 6-13 day babies was evaluated and a different cutoff was introduced. The high risk 6-13 day cutoff was set to the same percentile as the cutoff percentile (i.e., 0.15th) for the 1-6 day babies	6-13 day high risk cutoff introduced and reduced from 8.0 to 5.0 µmol/L/hr
06/04/2014	8µmol/L/hr	GLA percentile values (1st and median) were trending lower resulting in a higher retest rate. The trend was attributed to an increase in average temperature which causes a reduction in activity and a higher retest rate.	Decrease from 8.0 to 7.0 µmol/L/hr
6/23/2014	3.7 µmol/L/hr	14+ day cutoff was lowered to account for the lowering in the 1-6 day cutoff percentile due to weather related activity decrease	Decreased from 3.7 to 3.0 µmol/L/hr

* This false positive rate describes the false positive rate of MSPHL's screening procedure and not the false positive rate for the device as FDA has calculated it (please refer above to the section "estimated false positive rates" for a discussion on how these differ).

To explore the impact of the changes to the cutoffs during the pivotal study, Baebies provided a table summarizing the screening performance of the device for the entire pivotal phase of the study applying the final cutoffs used by MSPHL. The screening results are summarized in the table below:

GLA - Pivotal Phase						
		1 screen	2 screens	3 screens	4+ screens	Total
	Newborns	93,124	9,340	2,253	372	105,089
1st Test	All screens with first result above borderline	92,405	8,913	2,032	320	103,670
	At least one screen with first result below borderline	719	427	221	52	1,419
Average of all Tests	All screens w/ avg. above high risk	643	388	204	50	1,285
	At least one screen w/ avg. below high risk	76	39	17	2	134
	Referred	46	8	1	-	55
	Not Referred	31	31	15	2	79
	Referral Summary	1 screen	2 screens	3 screens	4+ screens	Total
Referred Sample Summary	Fabry - True Positive	24	4			28
	Fabry - Normal, False Positive	18	4	1		23
	Fabry - Refused	3				3
	Fabry - Moved	1				1
	Not Referred Summary	1 screen	2 screens	3 screens	4+ screens	Total
Samples Not Referred Summary	Prior Sample From Newborn Above Cutoff		22	14	1	37
	Other Assay Below Borderline	13	2			15
	Later Sample From Newborn Above Cutoff		7	1	1	9
	Outliers Excluded	9				9
	Transfused	5				5
	Different Cutoff Applied					0
	Spot Variability	1				1
	Retrospectively Referred					0
	Age Related Enzyme Decrease	1				1
	Contaminated Sample	1				1
	Multiple Reasons	1				1
Performance Summary	Total Presumed Normal	93,079	9,332	2,251	372	105,034
	Total Presumed Affected	46	8	1	-	55
	True Positives	24	4	0	0	28
	False Positives	18	4	1	-	23
	False Positive Rate (FPR)	0.019%	0.043%	0.044%	0.000%	0.022%

The impact of the changes to the cutoffs is that 66 babies that were considered high risk for Fabry (based on the cutoffs in use by MSPHL at the time these babies were screened) would no longer be considered high risk (i.e., during routine screening, 200 babies were considered high risk for Fabry based on the test result and using the final cutoffs, 134 babies were considered high risk for Fabry). The false positive rate for MSPHL's screening procedure is consistent at 0.022% (compared to 0.025%) and the false positive rate of the test (i.e., FDA's analysis) also remains consistent at 0.097% using the final cutoffs compared to 0.16% based on the cutoffs used at testing. Of note, among the 66 babies that are no longer high risk for Fabry are two confirmed positive Fabry babies. Another baby diagnosed during the pilot phase of the study is also no longer considered high risk for Fabry when the final cutoffs are applied to the original test results.

A summary of the screening performance of the GLA assay for all babies screened during the retrospectively-defined pivotal phase of the study using the final cutoffs is provided below. This table lists the total number of test screen negative babies (babies that were presumed normal following the initial testing and any repeat testing). Test screen positive is the number of babies that upon re-testing were considered high risk for Fabry (i.e., below the high risk cutoff). The table also includes number of test screen positive babies who were referred, those who were not referred, and those confirmed positive for the disease.

Screening performance GLA assay	Total number of newborns				
	1 screen	2 screens	3 screens	4+ screens	Total
Test screen negative	93,048	9,301	2,236	370	104,955
Test screen positive	76	39	17	2	134
Referred	46	8	1	0	55
Not referred	31	31	15	2	79
True positives	24	4	0	0	28

FDA would like the panel's input on whether the analysis of the clinical study should use the cutoffs used when testing each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff – cutoff period 8). This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

Analytical Performance of the GLA Assay

During our review of the analytical performance data provided in support of the GLA assay we noted the following:

Precision: Baebies performed a study to estimate the imprecision of the GLA assay using 4 analyzers, 3 reagent lots and performed testing during 21 non-consecutive days, with 2 runs per day and 2 dried blood spot (DBS) punches of each specimen per run. Baebies tested a total of 336 replicate results for each specimen. In this study 11 samples were identified as invalid and one was identified as an outlier. The results of the study are summarized below (please note that the outliers are included in this analysis). In the following table the repeatability (or within-run precision) estimate includes the instrument run as the component (or condition) of variability and the reproducibility estimate includes the instrument run, reagent lot, instrument and day as the components (or conditions) of variability. As a reminder, the final cutoffs during the pivotal study were 3.0, 5.0, 7.0 and 9.0 µmol/L/h.

Mean µmol/L/h	N	Repeatability (%CV)	Between Lot (%CV)	Between Instrument (%CV)	Between Day (%CV)	Reproducibility (%CV)
6.94	331	15.4%	5.6%	0%	1.7%	16.3%
9.80	334	10.4%	6.4%	0%	2.9%	13.6%
15.32	335	7.7%	7.2%	0%	3.1%	11.5%
28.76	334	7.8%	4.8%	0%	0.0%	9.4%
52.66	335	8.8%	4.1%	1.3%	2.7%	10.6%

The coefficient of variability (CV) was calculated for each imprecision estimate. The CV is calculated by dividing the standard deviation by the mean of the sample and describes the extent of variability in relation to the mean of the sample. The highlighted sections in the table above denote the imprecision of the GLA assay at the approximate concentrations of the cutoffs (high risk and borderline) used in the pivotal study (3.0, 5.0, 7.0. and 9.0 µmol/L/h).

Detection Limits: Baebies evaluated the detection limits of this assay following a recognized guideline[10] using 3 lots of their reagent and estimated the following detection limits:

- The limit of the blank (LoB) was defined as the highest analyte concentrations expected to be found when replicates of a sample containing no analyte are tested with 95% confidence. This is often a way of determining what concentration(s) the assay cannot distinguish from "noise." Baebies estimated that the LoB of the GLA assay was 1.96 µmol/L/h.

- The limit of detection (LoD) was defined as the lowest analyte concentration likely to be reliably distinguished from a blank sample with 95% confidence. The LoD of the GLA assay was determined to be 3.18 µmol/L/h (note the high-risk cutoff for the assay for babies 14+ days old was 3.0 µmol/L/h)

- The limit of quantification was defined as the lowest concentration where the total imprecision was ≤ 1.5 µmol/L/h or 20% CV whichever is greater. Baebies estimated that the LoQ for the GLA assay was

4.88 µmol/L/h which was the concentration where the imprecision was less than 1.5 µmol/L/h (and the CV could be as high as 31%). FDA estimates that the LoQ based on an imprecision goal of 20% CV (which is the typical imprecision goal for the LoQ of quantitative assays) is approximately 8 µmol/L/h (note all three high-risk cutoffs for the assay are ≤ 7.0 µmol/L/h).

Outliers: Baebies identified several statistical outliers in their analytical studies in support of the GLA assay (and all assays). As discussed above, MSPHL also identified "visual outliers" during the clinical study when clinical samples needed to be tested in multiple replicates.

Stability of the GLA samples (transport stability): Baebies performed a study showing the impact of shipping the DBS for up to 5 days at the following ambient conditions: 10ºC and 20% relative humidity (RH), 10ºC and 80% RH, 45ºC and 20% RH, 45ºC and 80% RH and 25ºC and 50% RH.

The impact on the test result is summarized in the table below:

Average GLA enzymatic activity for DBS samples before and after simulated transport

Condition	Day	Sample 3 Concentration µmol/L/hr		Sample 5 Concentration µmol/L/hr		Sample 10 Concentration µmol/L/hr		Sample 15 Concentration µmol/L/hr	
		before	after	before	after	before	after	before	after
10C, 20%RH	1	10.8	10.1	13.0	13.6	31.7	32.0	65.3	60.8
	3	10.9	10.8	13.6	14.7	31.1	31.7	61.1	56.2
	5	13.0	12.4	15.3	15.3	33.3	33.3	68.2	64.8
10C, 80%RH	1	12.4	12.8	14.8	14.9	36.1	33.5	70.5	67.7
	3	10.7	10.0	13.1	13.1	31.5	30.5	62.6	57.0
	5	11.7	10.4	14.8	14.9	33.0	30.7	65.9	58.0
25C, 50%RH	1	14.7	12.9	16.2	14.7	37.3	34.7	70.7	62.2
	3	12.4	12.0	15.5	13.8	35.3	28.9	65.8	56.6
	5	12.3	11.1	15.4	12.8	33.6	27.5	64.6	53.6
45C, 20%RH	1	11.7	8.8	14.5	10.3	33.1	20.9	67.6	37.8
	3	11.4	8.1	14.5	8.5	32.0	16.0	66.4	26.6
	5	12.0	8.2	15.2	8.6	33.2	14.6	62.4	21.2
45C, 80%RH	1	12.6	8.2	15.1	8.9	35.8	16.5	69.1	30.4
	3	12.8	5.8	15.6	6.3	34.9	8.7	68.8	11.0
	5	10.5	4.3	13.1	4.3	31.4	5.0	63.5	6.3

The samples above presented in red font are outside the acceptance criteria for acceptable stability defined by Baebies (i.e., recovery values <85% and >115% for enzyme concentrations greater than ≈6 µmol/L/hr were considered to be significantly impacted). The highlighted fields show samples with targeted concentrations set to represent normal samples close to the borderline cutoff (S3), low normal (S5) and high normal samples (S10 and S15) that upon transport at 45ºC and 20% and 80% relative humidity fall either below the borderline cutoff or some high risk cutoffs (borderline cutoff set to 5 to 10 µmol/L/hr and high risk cutoff ranged from 3.0 to 8 µmol/L/hr for this assay). These shipping conditions might be experienced in certain regions of the United States such as in the southern states especially if shipped at ambient conditions.

This study demonstrates that the enzyme activity will decrease when exposed to high temperature and/or humidity, and should be taken into account in certain geographic areas/seasons. This degradation could lead to

false positive results. FDA is asking for panel input on whether the panel is aware of any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity as result of standard shipping conditions.

Analytical Performance of the Assay at the Cutoffs

In summary:

- High risk cutoffs used during the course of the study (i.e., 3.7 µmol/L/h) were set below the LoQ of the assay (i.e., 4.88 µmol/L/h). The final high risk cutoff used for babies greater than 14 days old at the time the screen was collected (i.e., 3.0 µmol/L/h) was set below the LoD of the assay (i.e., 3.17 µmol/L/h. However, when the high risk cutoff is used, there were always multiple test results from the same DBS (any test result from previous screens or subsequent screens) available from each baby for analysis and interpretation.

- When samples from babies confirmed positive for Fabry were retested by MSPHL as part of their quality assurance, between 4.6 and 6.3% of the results fell above the borderline cutoff and between 21.4 and 22.5% of the results fell above the high risk cutoff.

- There were more outliers than is typical for assays of this type (statistical outliers during the analytical studies and visual outliers during the clinical study).

- Transport at certain ambient conditions resulted in decreased enzyme activity. We are aware that laboratories have reported seasonal variation with other analyte levels.

For this clinical use, newborn screening test manufacturers typically provide clinical data in the package insert that is analyzed with cutoffs not only set so that no known positive babies are incorrectly categorized (i.e., presumed normal instead of high risk for the screened condition and vice versa) but the cutoffs are also analytically valid (e.g., precise, reproducible). The rationale for this is so that laboratories can clearly understand the clinical performance of a screening assay set to not miss any known true positive babies with reliable cutoffs (precise, accurate, reproducible).

To summarize FDA's questions for the panel, we are seeking input on:

- Whether the Instructions for Use should utilize Baebies' false positive rate estimate (after the laboratory's risk analysis), the false positive rate estimate based on the test alone, or another specified analysis?

- Whether the panel has a recommendation on how to estimate the false negative rate of this device?

- Whether the analysis of the clinical study should use the cutoffs used to test each baby during the study, or the final cutoff (i.e., a retrospective analysis of the data using the final cutoff)? This input would guide FDA on what clinical performance characteristics of this device would be provided in the Instructions for use.

- Whether performance characteristics for age-specific cutoffs, when used, should be provided in the Instructions for use?

- Whether there are any measures that Baebies can recommend in their Instructions for Use to mitigate the impact on the enzyme activity standard shipping conditions, including high temperature and humidity?

- Whether the analytical performance of the assays at the cutoffs (e.g., precision, detection limits, outliers, performance of confirmed positive samples upon retesting) is sufficient for safe and effective use (i.e., does the panel consider that Baebies has demonstrated adequate analytical validity of the assays)?

V. Summary

As described above, Baebies conducted a large, prospective clinical study to assess the performance of the LSD assays on their Seeker System. This study provided a large volume of data, but due to multiple assay changes and complex clinical risk assessments performed as part of the study, it is difficult to understand how to assess the performance of the assay to determine whether it can be found safe and effective for its intended use. In addition, there are several potential analytical performance questions raised by the data submitted in support of this assay. FDA seeks the Panel's assistance in interpreting the data from these clinical and analytical studies to identify whether there are clinical or analytical concerns that should be addressed prior to regulatory approval. In addition, if the Panel believes the data can be adequate to support marketing authorization, FDA seeks advice on the information about clinical performance that should be included in the labelling so that public health laboratories have access to adequate instructions for use to assist them in safely implementing this assay in their laboratories.

VI. Panel Questions

FDA wishes to get input from the clinical community, via our Advisory Panel, to determine whether there are any concerns with the clinical and/or analytical performance of the Baebies Seeker System and also to get the panel's input on questions FDA has regarding statistical analysis of the clinical study and the relevant information that should be included in the Instructions for Use of the test system.

We have the following discussion questions for the panel to address during the Advisory Committee Meeting:

1. Typically, all babies that are determined to be high risk (i.e., for the Seeker System, a test result below the high risk cutoff) by a newborn screening test are presumed positive; in the statistical analysis of test performance these presumed positive results are determined to be either true positives as determined by clinical diagnosis or false positives. The pivotal study presented here used a risk analysis to determine those babies that should be referred for further diagnostic testing. Given that (1) for this pivotal study there is no follow-up information (i.e., diagnostic testing or clinical diagnosis) on the babies with presumed positive results that were not referred because of the assessment of the newborn's test results via the risk analysis (i.e., no clinical truth) and (2) since other laboratories may develop a different risk analysis or not use a risk analysis when using this device:

    a. Does the panel have a recommendation on how to calculate the false positive rate of this device?

    b. Similarly, does the panel have a recommendation on how to estimate the false negative rate of this device?

c. If an adequate estimation of the false positive and false negative result rate can be made based on this study, does the panel have concerns about the false positive and false negative rates observed in this study?

d. The risk analysis that MSPHL used in this study would be difficult to incorporate into the device and include in the Instructions for Use. Should the clinical risk analysis that was used in the study be included in the device, and if so, how?

2. It is unclear to FDA how the data should be analyzed and interpreted with respect to cutoffs.

   a. Should the analysis of the clinical study use the cutoffs used to test each baby during the study, the final cutoffs (i.e., a retrospective analysis of the data using the final cutoff), or another method? If this device is authorized for marketing, this input would guide FDA on the clinical performance characteristics of this device that would be described in the Instructions for use.

   b. Baebies did not provide screening performance estimates separately by the age of the baby at the time of screening (i.e., 1 to 6 days old, 7 to 13 days old and greater than 14 days old). Since different cutoffs were used for the GAA, GBA and GLA assays depending on the age of the newborn when the screen was performed, should the performance be provided by age at the time of screening in the final analysis?

   c. Based on the panel's recommendations for 2a and 2b, what information should be included in the device's instructions for use to guide the use of this test by other laboratories?

3. FDA has questions about the analytical performance of the assays at the cutoffs (e.g., precision, detection limits, outliers, performance of confirmed positive samples upon retesting) and whether that performance is adequate to ensure acceptable clinical test performance. Does the panel have any specific concerns with the analytical performance of the assays for each of the following? If so, please describe these concerns.

   a. The precision of the assays around the cutoffs?

   b. Typically the performance goal for the LoQ of quantitative assays, when defined based on imprecision alone (because there is no reference available to establish trueness), is the lowest concentration where the imprecision is less than or equal to 20% CV. Baebies defined the LoQ as the lowest concentration where the SD is less than or equal to 1.5 µmol/L/h. Depending on the assay, setting the LoQ in this fashion results in imprecision ranging from 31% CV (GLA assay) to 81% CV (GBA assay). Does the panel have any input regarding the appropriate performance goals for the limit of quantitation of these four assays (e.g., 1.5 SD, 20% CV, or other)?

   c. The presence of outliers in the analytical and clinical studies?

   d. The variation of the test result upon repeated measurements of the same samples?

4. Regarding sample instability, is the panel aware of any measures that Baebies can recommend in their Instructions for Use to mitigate loss of enzyme activity as result of standard shipping conditions, including high temperature and humidity?

5. Based on the information presented about the clinical and analytical data of the Seeker System, please discuss whether the benefits from the use of the Seeker System outweigh the risks of its use in the intended use population, and why?

VII. References

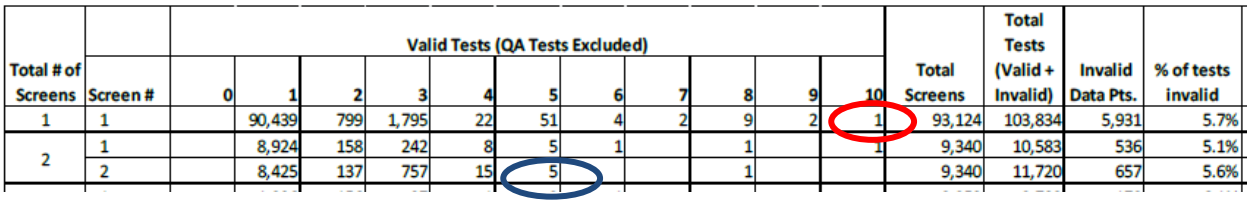
1. Bruni S, Loshchi L, Incerti C, et al. Update on treatment of lysosomal storage diseases. Acta Myologica 2007; 26: 87.
2. Wynn R. Mucopolysaccharidoses: Clinical features and diagnosis. In: Hahn S, Patterson MC, ed. Uptodate. Waltham, Mass.: Uptodate ; 2015.
3. Hughes D. Gaucher disease: Pathogenesis, clinical manifestations, and diagnosis. In: Hahn D, ed. Uptodate. Waltham, Mass.: Uptodate; 2015.
4. Merritt JL. Lysosomal acid alpha-glucosidase deficiency (Pompe disease, glycogen storage disease II, acid maltase deficiency). In: Hahn S, ed. Waltham, Mass.: Uptodate; 2015.
5. Mauer M, Kopp J. Clinical features and diagnosis of Fabry disease. In: Curhan G, ed. Uptodate. Waltham, Mass.: Uptodate; 2015.
6. Wang RY, Bodamer OA, Watson MS, et al. ACMG Standards and Guidelines-Lysosomal storage diseases: Diagnositic confirmation and management of presymptomatic individuals. Genetics in Medicine 2011; 13: 457.
7. Dasouki M, Jawdat O, Almadhoun O, et al. Pompe Disease: Literature Review and Case Series. Neurologic Clinics 2014; 32: 751.
8. Mauer M, Kopp J. Treatment of Fabry disease. In: Curhan G, Murphy , ed. Uptodate. Waltham, Mass.: Uptodate; 2015.
9. Branton MH, Schiffmann R, Sabnis SG, et al. Natural history of Fabry renal disease: influence of alpha-galactosidase A activity and genetic mutations on clinical course. Medicine (Baltimore) 2002; 81:122.
10. Clinical and Laboratory Standards Institute (CLSI). Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline—Second Edition. CLSI document EP17-A2. Clinical and Laboratory Standards Institute. Wayne, PA, 2012.

VIII. Appendix

The table below summarizes all testing peformed during the pivotal phase of the study.

Total # of Screens	Screen #	Valid Tests (QA Tests Excluded) 0	1	2	3	4	5	6	7	8	9	10	Total Screens	Total Tests (Valid + Invalid)	Invalid Data Pts.	% of tests invalid
1	1		90,439	799	1,795	22	51	4	2	9	2	1	93,124	103,834	5,931	5.7%
2	1		8,924	158	242	8	5	1		1		1	9,340	10,583	536	5.1%
	2		8,425	137	757	15	5			1			9,340	11,720	657	5.6%
3	1		1,996	156	97	1	2	1					2,253	2,789	170	6.1%
	2		2,038	69	139	4	2			1			2,253	2,799	172	6.1%
	3		1,805	38	398	10	2						2,253	3,317	192	5.8%
4	1		273	25	18	1							317	405	24	5.9%
	2		276	17	24								317	400	18	4.5%
	3		278	8	31								317	407	20	4.9%
	4		241	10	66								317	495	36	7.3%
5	1		38	5	2								45	55	1	1.8%
	2		38	2	5								45	61	4	6.6%
	3		35	1	8		1						45	74	8	10.8%
	4		36	2	7								45	65	4	6.2%
	5		38		7								45	68	9	13.2%
6	1		7			1							8	11	0	0.0%
	2		8										8	8	0	0.0%
	3		7		1								8	11	1	9.1%
	4		7		1								8	10	0	0.0%
	5		7	1									8	9	0	0.0%
	6		8										8	10	2	20.0%
7	1		1	1									2	3	0	0.0%
	2		2										2	2	0	0.0%
	3		2										2	2	0	0.0%
	4		1		1								2	4	0	0.0%
	5		1		1								2	4	0	0.0%
	6		1		1								2	4	0	0.0%
	7		1	1									2	3	0	0.0%
	Grand Tota	0	114,933	1,430	3,601	62	68	6	2	12	2	2	120,118	137,153	7,785	5.7%
	% of samples requiring	0.0%	95.7%	1.2%	3.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%				

In the table, the term “screen” describes obtaining a dry blood spot card with multiple blood spots from a baby (first screen is the initial card obtained; second screen is when a second card is requested and obtained, etc.). The term “test” refers to a punch taken from the DBS and analyzed. For example the circles below provide the following information:



Total # of Screens	Screen #	Valid Tests (QA Tests Excluded) 0	1	2	3	4	5	6	7	8	9	10	Total Screens	Total Tests (Valid + Invalid)	Invalid Data Pts.	% of tests invalid
1	1		90,439	799	1,795	22	51	4	2	9	2	1	93,124	103,834	5,931	5.7%
2	1		8,924	158	242	8	5	1		1		1	9,340	10,583	536	5.1%
	2		8,425	137	757	15	5			1			9,340	11,720	657	5.6%

The red circle indicates that 1 baby needed 10 tests from the DBS card obtained during the first screen in order to be categorized. The blue circle indicates that 5 babies needed 5 tests from the second screen (in addition to any testing performed for these babies using the DBS card obtained in the first screen) in order to be categorized. The table also shows that there were 93,124 babies with one valid screen. To categorize these babies, MSPHL performed 103,834 tests (5,931 of these tests were invalid). The table also shows that there were 9,340 babies with 2 valid screens during the pivotal phase. To categorize these babies, MSPHL performed 10,583 tests using the DBS card from the first screen and 11,720 tests from the DBS card from the second screen (for a total of 22,303 tests of which 1,193 (or 536 + 657) were invalid.