JIMD Reports
DOI 10.1007/8904_2013_266

INVITED ARTICLES

Systematic Data Collection to Inform Policy Decisions: Integration of the Region 4 Stork (R4S) Collaborative Newborn Screening Database to Improve MS/MS Newborn Screening in Washington State

Ashleigh Fleischman · John D. Thompson · Mike Glass

Received: 01 June 2013 / Revised: 10 September 2013 / Accepted: 13 September 2013 / Published online: 5 November 2013

Abstract In the past 50 years, newborn screening (NBS) has grown significantly in the breadth of screening programs and the number of conditions tested for each baby. The adaptation of tandem mass spectrometry (MS/MS) technology to detect inherited metabolic diseases is arguably one of the most impactful advancements in NBS testing. The addition of new conditions to the screening panel and the rarity of these conditions pose challenges for NBS program development, improvement, and evaluation. The Region 4 Stork (R4S) project is an international collaborative NBS database and a resource for programs across the world to overcome these challenges. By pooling true-positive case and laboratory testing data, the R4S database provides insight into complex MS/MS profiles for these rare conditions. The Washington State NBS Program is integrating aspects of the R4S web application and utilizing R4S resources to examine current protocols, identify improvements, implement changes, and review outcomes. Washington uses R4S resources to choose informative analytes and evaluate cutoffs. The program also examines the performance of R4S tools that are designed to aid in evaluating a baby's MS/MS screening results. This article documents these efforts in utilizing a subset of the R4S tools to improve their program, demonstrating the flexibility of the application. Other NBS programs can use the knowledge Washington has gained to strengthen their ability to correctly identify babies with metabolic disorders and mitigate the impact of screening on babies and their families.

Communicated by: Rodney Pollitt, PhD

Competing interests: None declared

A. Fleischman (✉) · J.D. Thompson · M. Glass
Washington State Department of Health, Office of Newborn Screening, 1610 NE 150th Street,
Shoreline, WA 98155, USA
e-mail: Ashleigh.Fleischman@doh.wa.gov

Introduction

The year 2013 marks the 50th anniversary of newborn screening (NBS), which began with the screening of phenylketonuria (PKU) in two US states. Since its inception, NBS has progressed and expanded, and is hailed as one of the greatest achievements in public health. Over the past 50 years, NBS has been implemented across most of the world and all babies born in the USA and in many other countries are now screened at birth for numerous dangerous or life threatening conditions. The adaptation of tandem mass spectrometry (MS/MS) to newborn screening during the 1990s allows NBS labs to screen for multiple amino acid, fatty acid, and organic acid metabolism disorders from one dried blood spot punch. Conditions on the MS/MS newborn screening panel are very rare, and for many of them, a NBS program may only see one true-positive case every few years. The rarity of true cases poses challenges for program development, improvement, and evaluation, particularly for smaller NBS programs. The Region 4 Stork (R4S) project was initiated in 2005 to pool information and generate data for NBS programs to improve MS/MS screening by (a) achieving uniformity of MS/MS testing panels, (b) improving analytical performance, and (c) reducing false-positive and false-negative screening results (McHugh et al. 2011). The project was initiated by the Region 4 Genetics Collaborative and funded through grants from the federal Health Resources and Services Administration (http://www.region4genetics.org/). In 2008 the R4S website was launched, allowing participant

NBS programs to upload information into the database and have live up-to-date access to a number of helpful screening tools developed by the R4S project utilizing the collaborative data. The R4S website was recently described in detail (McHugh et al. 2011). As of May 2013, the database contains NBS results for over 15,000 true-positive cases from 49 US states and 42 countries (https://www.nbstrn.org/research-tools/lab-performance-database). An average of five new cases are added to R4S every day; this is more true-positive cases than some NBS programs will find in a year. By compiling case data and laboratory results from across the world, the database provides insight into the complex MS/MS profiles for the rare conditions on the NBS panel. The comprehensive R4S website can be used by NBS programs to examine condition profiles, identify informative markers, examine site-specific cutoffs, and improve test sensitivity and specificity using the R4S-generated tools. R4S contains a multitude of tools that can be customized and applied in hundreds of ways to NBS programs. The Washington State (WA) NBS Program participated in the website training program and continues to actively contribute to the R4S database. This article highlights the specific ways by which WA NBS has employed R4S functions to strengthen screening operations.

Washington Newborn Screening

In the Washington State Newborn Screening Program, the laboratory and follow-up teams reside in the same building, in a single administrative unit and work together to develop screening methods for each condition on the required panel. It is the responsibility of the follow-up team to review and evaluate screening algorithms for continual program improvement. The team stays abreast of new techniques and advancements in the screening community and several staff members participated in the training program for R4S hosted by the Region 4 Genetics Collaborative. The weeklong course reviewed the R4S database and taught users how to use the R4S website and tools. The WA NBS program has taken the knowledge gained in this course and applied many of the site functions in program evaluation and improvement efforts.

Across the world, NBS programs differ on screening procedures, cutoff algorithms, and condition panels. In Washington State, the State Board of Health is the governing body which oversees the screening panel and determines which conditions are required for screening based on five criteria: (1) prevention potential and medical rationale, (2) treatment availability, (3) public health rationale, (4) available technology, and (5) cost-benefit/cost-effectiveness. In 2008, Washington underwent a rigorous NBS advisory committee review with the State Board of Health to expand the panel using MS/MS technology. Currently, Washington is screening for 27 of the conditions on the US federal government's Recommended Uniform Screening Panel (RUSP); 19 of these are tested using MS/MS technology. This is not an extensive panel of conditions detectable through MS/MS screening, and in order to limit incidental findings (conditions not approved by the State Board of Health), MS/MS testing is done using multiple reaction monitoring (MRM) which only measures specific analytes defined by the user. This differs from many NBS labs that use a full MS/MS scan which measures all analytes within a specified range, resulting in identification of many disorders that are not on Washington's panel. In 2008, the analytes selected for the MS/MS panel were determined using information available at the time from R4S and other laboratories. With the progression of the R4S database as a resource of true-positive cases and their associated laboratory findings, the profiles of these rare conditions are now better understood.

Analyte Selection and Review

With the limitations of MRM, it is critical that each analyte provide useful information. The tools within R4S provide a quick and easy way to identify informative markers (analytes and analyte ratios) for all the NBS conditions. The best screening markers are out of range in the affected population and normal in the unaffected population, with good separation between the two. Rarely does this happen for any screening test, and for most conditions, a combination of markers is used to help tease out the differences between these two populations. Using the R4S tools, the MS/MS panel in WA can be reviewed and modified as necessary to ensure its full potential to identify the required conditions. This can be done quickly and efficiently.

To optimize the MS/MS panel, WA NBS needed to identify the most informative markers for each condition on the panel. This was done using two of the R4S Project Tools, one which examines analytes by condition and one which examines conditions by analytes. The first tool, called Plots by Condition, generates box plots of the disease and normal ranges for markers associated with a specific condition. This can be done for all acylcarnitines, acylcarnitine ratios, amino acids, and amino acid ratios measured by MS/MS. Plots can be filtered to view only those analytes considered informative for a particular condition (categorized by the median of the affected population being outside the 90th percentile of the unaffected population). Markers where there is no overlap between the two populations are the most informative and likely have clinical utility for

interpretation of screening results. Individual plots can be produced for every condition in the R4S database. The Plots by Condition tool can quickly identify markers that should be included on the MS/MS MRM panel.

Conversely, a tool called Plot by Marker (available for every marker) generates box plots of the disease and normal ranges for conditions associated with a specific marker. A marker is considered informative for a condition if the median of the affected population is outside the normal range. Conditions where the marker is most informative have no overlap between the two populations. Once again, these plots quickly identify conditions in which the marker is out of range and users can easily identify the potential influence of each additional analyte or analyte ratio. The information from R4S can be used to identify the markers which will make the most impact and limit incidental findings for programs using MRM. Washington undertook a careful review, comparing the current MS/MS panel with the identified informative markers. The review resulted in the addition of nine new analytes and nine new ratios, impacting nine conditions on the screening panel. Twelve of the 18 new markers had no or very little overlap between the affected and unaffected populations.

Cutoff Determinations

The addition of new conditions to a screening panel can be a long and arduous process. Often there is only limited knowledge of the natural progression of the disease, and data for disease and normal ranges may be based on only a small nonrepresentative population. In the past, Washington has mainly used experiences from other programs when determining initial cutoff algorithms for new screening tests. When the expansion of MS/MS screening occurred in 2008, preliminary R4S data assisted in cutoff determinations. Now, R4S has a variety of tools that can be used to review, compare, and improve cutoffs. The Analyte Comparison tool visually displays the disease and normal ranges and cutoff values from all participants for individual analytes. The users can quickly compare their normal range and cutoffs to those of other programs. They can also see if their cutoff value is within a recommended target range and if it overlaps the normal range of their population. This tool has been helpful in evaluating the cutoffs historically used in Washington.

The Normal Percentiles Comparison tool can be used to compare program normal ranges to those in the database, identifying analytes in which they differ from the R4S population. Differences may be attributable to testing methods, population, or errors in data. If program-specific normal ranges differ vastly from the R4S population, recommended cutoff ranges may not be appropriate for that particular analyte. If the normal ranges are comparable, a Score Card provides recommended cutoff values, with percentiles for the normal population, participant cutoffs, and true-positive disease ranges. The user can then choose a cutoff value using the available percentile ranges. In Washington, cutoffs are reviewed periodically using these tools to evaluate their effectiveness. For the 18 new analytes and ratios where Washington-specific normal ranges were not available, initial cutoffs were chosen using the target range from the Score Card. After normal range data can be collected, temporary cutoffs will be reviewed to ensure they are appropriate for the Washington population. R4S provides a robust data set to support decisions made in program review of cutoffs and the addition of new analytes to a screening panel.

Post-Analytical Tools

It is well known in the screening community that using cutoffs is not a perfect method for detecting conditions; there is an art to the science of balancing false-positives and false-negatives. Programs try to mitigate errors by using multiple markers, demographics, and clinical information to make determinations on screening results. R4S has developed Post-Analytical (PA) tools to assist programs in making these decisions, and in the case of Minnesota's MS/MS panel, to eliminate cutoffs all together (Rinaldo 2013). A recent paper documents how the PA tools use multivariate pattern-recognition software to create case scores based on the MS/MS results uploaded into R4S, in particular the degree of overlap of informative analytes (Marquardt et al. 2012). As with many other programs, Washington is uneasy with eliminating cutoff algorithms, but is interested in finding ways to integrate the PA tools into the program as an additional resource in the evaluation of test results. A plan was developed to use the PA tools in parallel with the current cutoff scheme and regard them as a guide in decision-making processes. NBS programs may find the Tool Runner to be most effective because it can process batched NBS data directly from the MS/MS instruments, generating hundreds of scores instantly. Washington NBS is not routinely using this tool due to some logistical challenges but would like to increase its usage in the future.

Currently, WA NBS routinely uses the All Condition tool. Screening data is uploaded directly into the application and the All Condition tool runs every One Condition PA tool simultaneously. It then outputs an overview of each PA tool that has a summary of tools with informative results. Each One Condition tool contains useful information specific to that condition, including normal percentiles, disease ranges, and percent of overlap between the two.

The data is also provided in graphical form which is instantly helpful as the user can visually compare case analyte values against the normal and disease ranges. There is also a clear visual categorization of the informative markers. To assist in interpretation, each tool is programmed to generate a Case Score. These scores can be compared to other true-positive cases both in a Percentile Rank and in a graphical presentation. Interpretation Guidelines categorize the Case Scores into *not informative*, *possibly* disease, *likely* disease, and *very likely* disease using percentile distributions of true-positive cases. The Case Score interpretation is intended to provide guidance for follow-up programs. Some Case Score interpretations recommend further screening tests that are not available in most programs. These programs must decide how to proceed without the additional information provided by the recommended tests. For states with routine second screens, like Washington, the decision process may be even more complex. The question many programs ask is how do these Case Score interpretations translate into their population? What does *very likely* mean? Does every *very likely* translate into a true-positive case? Does every true-positive receive a score of *very likely*? Are some tools better than others at predicting the outcome? Washington NBS is collecting data to help answer these questions. Every abnormal specimen is run through the All Condition tool and the Case Score, Percentile Rank, and Interpretation are recorded for each one.

The PA tools have been run hundreds of times in Washington over the past few years, but the number of true-positive cases with MS/MS data available is relatively small for some conditions, inhibiting conclusive answers to the questions above. However, preliminary numbers are providing insight into the utility of some of the PA tools. In R4S, there are four One Condition tools where propionylcarnitine (C3) is considered the primary marker: (1) propionic acidemia (PROP), (2) methylmalonic acidemia-mutase or Cobalamin A and B deficiencies (MUT/Cbl A,B), (3) methylmalonic acidemia-Cobalamin C and D deficiencies (Cbl C,D), and (4) maternal Vitamin B12 deficiency (B12 Def (mat)). Table 1 depicts the distribution of Case Score interpretations for all specimens in Washington where the C3 was considered abnormal on the first screen and the baby was referred for diagnostic testing. Table 1 also includes the positive predictive value (PPV) of One Condition tool interpretations. The PPV of a *very likely* interpretation varied across the PA tools (28–100%), with the PROP tool having 100% PPV, but including only two cases. All true-positive C3 cases received an interpretation of *very likely*, with two exceptions. One received an interpretation of *possibly*; however, this was a case of maternal Vitamin B12 deficiency, which is considered an incidental finding of NBS. The other exception was an uncommon presentation of mild propionic acidemia: the baby was 6 months old when finally diagnosed, he was asymptomatic and only treated briefly. When combining *very likely*, *likely*, and *possibly* interpretations compared to a *not informative* interpretation, the PPV becomes more consistent across the different PA tools, ranging from 21% to 33%. If all four PA tools are combined, the sensitivity is 94% and the PPV is 25%. Determining the PPV for MS/MS analytes will allow NBS follow-up to provide this information at the time of referral and give the families a better idea of the likelihood that their baby is affected.

The Dual Scatter Plot PA tools can also be valuable assets to NBS programs. These tools compare two One Condition PA tools to determine if the profile is more consistent with one condition or the other. These can be used to separate serious conditions from heterozygotes (het) and milder forms of the condition. For example, the Dual Scatter Plot Guidelines for the VLCAD vs VLCAD(het) tool are: *VLCAD*, *VLCAD(het)*, *neither*, or *not informative* (meaning it could be either VLCAD deficiency or a VLCAD carrier). Data in Table 2 are for commonly used Dual Scatter Plots in Washington where specimens were abnormal on the first NBS and diagnostic testing was performed. The preliminary data for the Dual Scatter Plot tools show that when disease status was predicted (*VLCAD*, *MCAD*, or *PKU*) the tool was correct 100% of the time. However, not every true-positive case was as clear: some received scores in the uninformative category indicating the tool could not distinguish between the condition on the NBS panel and the heterozygous or mild state. Over the past few years, Washington has experienced some difficulty with C14:1 having a high false-positive rate for VLCAD deficiency: in the past year, there have been 51 babies with elevated C14:1 on the first screen. Forty-six of these had a C14:1 value under 1.0 (approximately 80% of babies with VLCAD have a C14:1 value greater than 1.0 on the NBS). Washington found that the vast majority of babies being referred were either VLCAD carriers or false-positive newborn screens (VLCAD and VCLAD carriers are diagnosed by DNA sequencing). In order to reduce the amount of unnecessary diagnostic testing, WA NBS has piloted a procedure using the VLCAD vs VLCAD(het) Dual Scatter Plot tool. When NBS results show an elevated C14:1 that is less than 1.0 with one or more secondary ratios in the normal range, specimens are run through the VLCAD vs VLCAD(het) Dual Scatter Plot tool. If the results are clearly in the *VLCAD(het)* range, the baby is not referred for diagnostic testing and a second NBS is requested. If the tool predicts either *VLCAD* or *not informative,* the baby is referred for diagnostic testing. There was only one true-positive mild VLCAD case where the Dual Scatter Plot indicated the results were *not informative*. Since the implementation of this protocol, 28

Table 1 Region 4 Stork (R4S) Case Score interpretations and positive predictive values (PPV) for babies in Washington State with elevated propionylcarnitine (C3) on the first newborn screen and diagnostic testing outcomes

Interpretation	Diagnostic outcome (*n*)		PPV (%)
PROP tool	**True-positive**	**Normal**	
Very likely	2	0	100
Possibly	0	4	0
Not informative	1	12	-
MUT/Cbl A,B tool			
Very likely	5	8	39
Likely	0	7	0
Possibly	0	1	0
Cbl C,D tool			
Very likely	3	5	38
Likely	0	5	0
Possibly	0	4	0
Not informative	0	2	-
B12 Def (mat) tool			
Very Likely	5	13	28
Likely	0	2	0
Possibly	1	1	50
All C3 PA tools	**True-positive**	**Normal**	
Very likely	15	26	37
Likely	0	14	0
Possibly	1	7	13
Not informative	1	14	-

PROP: propionic acidemia, *MUT/Cbl A,B*: methylmalonic acidemia (mutase or Cobalamin A and B deficiencies), *Cbl C,D*: methylmalonic acidemia (Cobalamin C and D deficiencies), *B12 Def (mat)*: maternal Vitamin B12 deficiency

Table 2 Region 4 Stork (R4S) Dual Scatter Plot interpretations for babies in Washington State with abnormal results on the first newborn screen and diagnostic testing outcomes

Interpretations	Diagnostic outcome (*n*)		
VLCAD vs VLCAD(het) tool	**VLCAD**	**VLCAD(het)**	**Normal**
VLCAD	1	0	0
VLCAD(het)	0	3	5
*Not informative**	1	10	1
Neither	0	0	2
MCAD vs MCAD(het) tool	**MCAD**	**MCAD(het)**	**Normal**
MCAD	3	0	0
MCAD(het)	0	0	1
PKU vs H-Phe tool	**PKU**	**H-Phe**	**Normal**
PKU	1	0	0
*Not informative***	3	4	0

very long-chain acyl-CoA dehydrogenase (VLCAD) deficiency vs very long-chain acyl-CoA dehydrogenase (VLCAD) heterozygote (het), medium-chain acyl-CoA dehydrogenase (MCAD) deficiency vs medium-chain acyl-CoA dehydrogenase (MCAD) heterozygote (het), pheylketonuria (PKU) vs hyperphenylalaninemia (H-Phe)

*Could be either VLCAD or VLCAD (het)

**Could be either PKU or H-Phe

babies and their families have avoided the fiscal and emotional expense of a diagnostic referral for VLCAD. These Dual Scatter Plot tools provide an opportunity for programs to adjust cutoffs to reduce false-positives, and to improve the PPV of their screens.

Limitations

While Washington has found many aspects of the R4S web application useful in program evaluation and improvement, there are some limitations to its utility. Currently there is no definition of a true-positive case: each participant uses different clinical measures to determine if a case is a true-positive newborn screen or not. Fortunately, every case that is uploaded into the system is reviewed by the curators for extreme values in an effort to reduce skewing of the results. With the pooling of data and the exclusion of outliers, misclassified results should have little effect on the tools and outcomes.

Some limitations only apply to a subset of the screening community. The MS/MS Portal is only validated for specimens less than 10 days of age at collection, indicating the tools may not be as useful for programs with routine second specimens collected outside the 10-day window, or for the NICU population (which generally receives three routine newborn screens in the USA). R4S has developed a second portal called MS/MS[2] in an effort to capture data specific to specimens collected at an older age. The R4S system was designed so that users can request portals, such as MS/MS[2], add conditions or ratios, and build their own PA tools. With only five participants and less than 400 true-positive cases, the MS/MS[2] Portal does not yet have the level of participation of the MS/MS Portal, but Washington plans to collaborate with other two-screen NBS programs to increase the utility of MS/MS[2].

Finally, there are limitations to the data available in Washington. The small number of true-positive cases is a barrier to complete validation of the PA tools available in R4S. Following the example set by R4S, collaboration with other NBS programs would increase the power of the validation study started in Washington, providing an opportunity to reach evidence-based conclusions on the utility of the Post-Analytical tools in a live NBS program.

Conclusion

Washington will continue to explore the integration of R4S into its screening program and is happy to serve as a resource for other programs who hope to do the same. When faced with the reality of screening for rare and deadly disorders in a climate where funding is scarce and public perception can be easily swayed, it is paramount for NBS programs to utilize available resources to ensure screening is performed quickly, efficiently, and reliably. R4S is a shining example of how teamwork within the screening community can provide improved outcomes for newborns across the globe. Not all of the R4S tools will be applicable to every NBS program; Washington State's experience using the tools to choose informative analytes, evaluate cutoffs, and collect longitudinal data on performance of the PA tools demonstrates that using a subset of individual R4S tools can strengthen a program's ability to correctly identify babies with metabolic disorders and mitigate the impact of screening on babies and their families.

Synopsis

Newborn screening programs can integrate tools from the Region 4 Stork (R4S) collaborative database to evaluate current protocols, identify improvements, and review outcomes of implemented changes. The power of a large pool of data for rare conditions makes R4S an excellent resource for newborn screening programs.

Compliance with Ethics Guidelines

Conflict of Interest

Ashleigh Fleischman, John D. Thompson, and Mike Glass declare that they have no conflicts of interest.

Animal Rights and Human Subjects

This article does not contain any studies with human or animal subjects performed by any of the authors.

Details of the Contributions of Individual Authors

Ashleigh Fleischman attended the Region 4 Stork MS/MS training and is assigned to follow-up for abnormal MS/MS conditions. She performed the review of the MS/MS panel, identified new markers to add to the panel, executed the cutoff review, proposed cutoff changes, conducted validation of the Post-Analytical tools including running specimens through the R4S tools, and recording results. Ashleigh was responsible for generating and reviewing the data and is the primary author of this article.

John D. Thompson is the supervisor for the follow-up group in Washington State's Newborn Screening program.

He attended the Region 4 Stork MS/MS training, collaborated and approved changes in the Washington MS/MS panel and cutoff algorithms, and oversaw the data collection and validation of the Post-Analytical tools. He served as an advisor for this article including planning, editing, and content review.

Mike Glass is the director of Washington's newborn screening program. He reviewed and approved all changes made in follow-up or lab protocols. He served as an advisor for this article including, planning, editing, and content review.

References

Marquardt G, Currier R, McHugh DMS et al (2012) Enhanced interpretation of newborn screening results without analyte cutoff values. Genet Med 14(7):648–655

McHugh DMS, Cameron CA, Abedenur JE et al (2011) Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. Genet Med 3(3):230–254

Rinaldo P (2013) The C. Ronald Scott lecture series in biochemical genetics. The impact, products and future applications of the Region 4 Stork (R4S) collaborative project. May 23 [lecture]. Seattle Children's Hospital, Seattle