

Enhanced interpretation of newborn screening results without analyte cutoff values

Gregg Marquardt, MSS¹, Robert Currier, PhD², David M.S. McHugh¹, Dimitar Gavrillov, MD, PhD¹, Mark J. Magera¹, Dietrich Matern, MD¹, Devin Oglesbee, PhD¹, Kimiyo Raymond, MD¹, Piero Rinaldo, MD, PhD¹, Emily H. Smith, PhD¹, Silvia Tortorelli, MD, PhD¹, Coleman T. Turgeon¹, Fred Lorey, PhD², Bridget Wilcken, MD³, Veronica Wiley, PhD³, Lawrence C. Greed, BSc⁴, Barry Lewis, MD⁴, François Boemer, PharmD, PhD⁵, Roland Schoos, PhD⁵, Sandrine Marie, PhD⁶, Marie-Françoise Vincent, MD, PhD⁶, Yuri Cleverthon Sica, MSc⁷, Mouseline Torquado Domingos⁷, Khalid Al-Thihli, MD⁸, Graham Sinclair, PhD⁸, Osama Y. Al-Dirbashi, PhD⁹, Pranesh Chakraborty, MD⁹, Mark Dymerski¹⁰, Cory Porter¹⁰, Adrienne Manning¹¹, Margretta R. Seashore, MD¹², Jonessy Quesada, MD¹³, Alejandra Reuben¹³, Petr Chrastina, MSc¹⁴, Petr Hornik, PhD¹⁴, Iman Atef Mandour, MD¹⁵, Sahar Abdel Atty Sharaf, MD¹⁵, Olaf Bodamer, MD, PhD¹⁶, Bonifacio Dy, MD¹⁷, Jasmin Torres¹⁷, Roberto Zori, MD¹⁸, David Cheillan, PhD¹⁹, Christine Vianey-Saban, PhD¹⁹, David Ludvigson²⁰, Adrya Stenbridge²¹, Jim Bonham, PhD²², Melanie Downing, MSc²², Yannis Dotsikas, PhD²³, Yannis L. Loukas, PhD²³, Vagelis Papakonstantinou, PhD²⁴, Georgios S.A. Zacharioudakis, PhD²⁴, Ákos Baráth, PhD²⁵, Eszter Karg, MD, PhD²⁵, Leifur Franzson, PhD²⁶, Jon J. Jonsson, MD, PhD²⁶, Nancy N. Breen²⁷, Barbara G. Lesko²⁷, Stanton L. Berberich, PhD²⁸, Kimberley Turner, RN²⁹, Margherita Ruoppolo, MD³⁰, Emanuela Scolamiero³⁰, Italo Antonozzi, MD³¹, Claudia Carducci, MS³¹, Ubaldo Caruso³², Michela Cassanello³², Giancarlo la Marca, Pharm Sc³³, Elisabetta Pasquini, MD³⁴, Iole Maria Di Gangi, PhD³⁵, Giuseppe Giordano, PhD³⁵, Marta Camilot, PhD³⁶, Francesca Teofoli³⁶, Shawn M. Manos, BS³⁷, Colleen K. Peterson, BS³⁷, Stephanie K. Mayfield Gibson, MD³⁸, Darrin W. Sevier³⁸, Soo-Youn Lee, MD, PhD³⁹, Hyung-Doo Park, MD, PhD³⁹, Issam Khneisser, MS⁴⁰, Phaidra Browning⁴¹, Fizza Gulamali-Majid, PhD⁴², Michael S. Watson, PhD⁴³, Roger B. Eaton, PhD⁴⁴, Inderneel Sahai, MD⁴⁴, Consuelo Ruiz⁴⁵, Rosario Torres⁴⁵, Mary A. Seeterlin, PhD⁴⁶, Eleanor L. Stanley⁴⁶, Amy Hietala⁴⁷, Mark McCann⁴⁷, Carlene Campbell⁴⁸, Patrick V. Hopkins⁴⁸, Monique G. de Sain-Van der Velden, PhD⁴⁹, Bert Elvers⁵⁰, Mark A. Morrissey, PhD⁵¹, Sherlykuty Sunny⁵¹, Detlef Knoll, MSc⁵², Dianne Webster, PhD⁵², Dianne M. Frazier, PhD⁵³, Julie D. McClure, MPH⁵³, David E. Sesser⁵⁴, Sharon A. Willis⁵⁴, Hugo Rocha, MSc⁵⁵, Laura Vilarinho, PhD⁵⁵, Catharine John, PhD⁵⁶, James Lim, PhD⁵⁶, S. Graham Caldwell⁵⁷, Kathy Tomashitis, MNS⁵⁷, Daisy E. Castiñeiras Ramos⁵⁸, Jose Angel Cocho de Juan, PhD⁵⁸, Inmaculada Rueda Fernández, MD⁵⁹, Raquel Yahyaoui Macías, MD, PhD⁵⁹, José María Egea-Mellado⁶⁰, Inmaculada González-Gallego, PhD⁶⁰, Carmen Delgado Pecellin, PhD⁶¹, Maria Sierra García-Valdecasas Bermejo, PhD⁶¹, Yin-Hsiu Chien, MD, PhD⁶², Wuh-Liang Hwu, MD, PhD⁶², Thomas Childs, MT(ASCP)⁶³, Christine D. McKeever⁶³, Tijen Tanyalcin, MD, PhD⁶⁴, Mahera Abdulrahman, MD, PhD⁶⁵, Cecilia Queijo, PhD⁶⁶, Aída Lemes, MD⁶⁶, Tim Davis⁶⁷, William Hoffman⁶⁷, Mei Baker, MD⁶⁸ and Gary L. Hoffman⁶⁸

¹Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, Minnesota, USA; ²California Department of Public Health, Richmond, California, USA; ³The Children's Hospital at Westmead, Sydney, New South Wales, Australia; ⁴Department of Clinical Biochemistry, Princess Margaret Hospital, Perth, Western Australia, Australia; ⁵Centre Hospitalier Universitaire de Liège, Liège, Belgium; ⁶Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Brussels, Belgium; ⁷Fundação Eucêmica de Proteção ao Excepcional, Curitiba, Brazil; ⁸Children's & Women's Health Center, Vancouver, British Columbia, Canada; ⁹Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; ¹⁰Colorado Department of Public Health and Environment, Denver, Colorado, USA; ¹¹Connecticut Department of Public Health Laboratory, Hartford, Connecticut, USA; ¹²Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; ¹³Hospital Nacional de Niños, San José, Costa Rica; ¹⁴Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; ¹⁵Cairo University Faculty of Medicine, Cairo, Egypt; ¹⁶Department of Human Genetics, University of Miami Miller School of Medicine, Miami, Florida, USA; ¹⁷Florida Newborn Screening Program, Jacksonville, Florida, USA; ¹⁸Department of Pediatrics, University of Florida, Gainesville, Florida, USA; ¹⁹Hospices Civils de Lyon, Lyon, France; ²⁰Georgia Department of Public Health, Atlanta, Georgia, USA; ²¹Department of Human Genetics, Emory University, Atlanta, Georgia, USA; ²²Sheffield Children's NHS Foundation Trust, Sheffield, UK; ²³School of Pharmacy, University of Athens, Athens, Greece; Submitted 31 August 2011; accepted 28 December 2011; advance online publication 16 February 2012. doi:10.1038/gim.2012.2

(Affiliations continue on the following page)

Purpose: To improve quality of newborn screening by tandem mass spectrometry with a novel approach made possible by the collaboration of 154 laboratories in 49 countries.

Methods: A database of 767,464 results from 12,721 cases affected with 60 conditions was used to build multivariate pattern recognition software that generates tools integrating multiple clinically significant results into a single score. This score is determined by the overlap between normal and disease ranges, penetration within the disease range, differences between conditions, and weighted correction factors.

Results: Ninety tools target either a single condition or the differential diagnosis between multiple conditions. Scores are expressed as the percentile rank among all cases with the same condition and

are compared to interpretation guidelines. Retrospective evaluation of past cases suggests that these tools could have avoided at least half of 279 false-positive outcomes caused by carrier status for fatty-acid oxidation disorders and could have prevented 88% of known false-negative events.

Conclusions: Application of this computational approach to raw data is independent from single analyte cutoff values. In Minnesota, the tools have been a major contributing factor to the sustained achievement of a false-positive rate below 0.1% and a positive predictive value above 60%.

Genet Med 2012;14(7):648–655

Key words: cutoff values; false-positive rate; inborn errors of metabolism; newborn screening; positive predictive value.

INTRODUCTION

The Regional Genetics and Newborn Screening Collaboratives funded by the Maternal and Child Health Bureau have been very successful in improving the newborn screening infrastructure of the United States. One of these initiatives has supported a project to hasten the implementation of newborn screening by tandem mass spectrometry (MS/MS)¹ and achieve uniformity of targets.² The importance of this endeavor is underscored by the recent inclusion of expanded newborn screening among the 10 great public health achievements of the past decade in the field of maternal and infant health.³

The specific objectives of the collaborative project are (i) to achieve consistency with the uniform panel adopted as the national standard by the Secretary of Health and Human Services⁴ and (ii) to improve analytical performance through the pursuit of the lowest achievable rates of false-positive and false-negative results.⁵ This project has grown to include 154 public health programs and private laboratories worldwide, leading to the publication of 8,255 disease ranges and 114 cutoff target ranges for amino acids, acylcarnitines, and related ratios.^{6,7}

We have developed multivariate pattern-recognition software designed to convert metabolic profiles into a composite score driven by the degree of overlap between normal population and disease range. Clinical relevance of a marker is reached when

the median of the disease range is outside the percentile limits of the normal population.⁶ A simultaneous assessment of multiple analytes is performed according to the degree of penetration within the respective disease range, expected differences between specific conditions, and proportionally weighted correction factors. This approach could represent a viable alternative to analyte cutoff values in the process of raw data interpretation, fostering their replacement with score-interpretation guidelines for a given condition.

MATERIALS AND METHODS

The Region 4 Stork MS/MS data project is a Web-based application developed using Microsoft.NET framework 3.5 and SQL Server 2008.⁶ The criteria for case definition are set by the local protocols of the individual participating sites and by overarching requirements that have been described previously.⁶ As of 15 December 2011, the MS/MS profiles of 12,077 patients affected with 60 metabolic disorders and of 644 heterozygote carriers for 12 conditions have been collected in this database. These profiles have served as the training set for the development of the postanalytical tools, and their number continues to expand. Since the beginning of 2009, an average of 5.2 new cases has been added per day (2008: 1,796 cases; 2009: 1,734 cases; 2010: 1,452 cases). The current population study translates to 767,408 discrete analyte concentrations and calculated

²⁴Neolab SA, Athens, Greece; ²⁵Department of Pediatrics, University of Szeged, Szeged, Hungary; ²⁶Department of Genetics and Molecular Medicine, Landspítali University of Iceland, Reykjavik, Iceland; ²⁷Indiana Newborn Screening Laboratory, Indianapolis, Indiana, USA; ²⁸State Hygienic Laboratory at the University of Iowa, Ankeny, Iowa, USA; ²⁹University of Iowa Children's Hospital, Iowa City, Iowa, USA; ³⁰CEINGE—Biotecnologie Avanzate, Università degli Studi di Napoli "Federico II," Naples, Italy; ³¹Sapienza University of Rome, Rome, Italy; ³²University Department of Pediatrics, G. Gaslini Institute, Genoa, Italy; ³³Department of Pharmacology, University of Florence, Florence, Italy; ³⁴Meyer Children's Hospital, Florence, Italy; ³⁵Department of Pediatrics, Università di Padova, Padua, Italy; ³⁶Azienda Ospedaliera Universitaria Integrata di Verona, Verona, Italy; ³⁷Kansas Department of Health & Environmental Laboratories, Topeka, Kansas, USA; ³⁸Kentucky Department for Public Health, Frankfort, Kentucky, USA; ³⁹Sungkyunkwan University School of Medicine, Seoul, South Korea; ⁴⁰Saint Joseph University, Beirut, Lebanon; ⁴¹Tulane University, New Orleans, Louisiana, USA; ⁴²Department of Health and Mental Hygiene, Baltimore, Maryland, USA; ⁴³American College of Medical Genetics, Bethesda, Maryland, USA; ⁴⁴New England Newborn Screening Program, Boston, Massachusetts, USA; ⁴⁵Universidad Autónoma de Nuevo León, Monterrey, Mexico; ⁴⁶Michigan Department of Community Health, Lansing, Michigan, USA; ⁴⁷Minnesota Department of Health, St. Paul, Minnesota, USA; ⁴⁸Missouri Public Health Laboratory, Jefferson City, Missouri, USA; ⁴⁹University Medical Center, Utrecht, The Netherlands; ⁵⁰National Institute for Public Health and the Environment, Bilthoven, The Netherlands; ⁵¹Wadsworth Center, New York State Department of Health, Albany, New York, USA; ⁵²LabPlus, Auckland Hospital, Auckland, New Zealand; ⁵³Division of Genetics and Metabolism, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina, USA; ⁵⁴Oregon State Public Health Laboratory, Hillsboro, Oregon, USA; ⁵⁵National Institute of Health Doutor Ricardo Jorge, Porto, Portugal; ⁵⁶KK Women's and Children's Hospital, Singapore, Singapore; ⁵⁷South Carolina Department of Health and Environmental Control, Columbia, South Carolina, USA; ⁵⁸Hospital Clínico Universitario, Santiago de Compostela, Spain; ⁵⁹Carlos Haya University Hospital, Málaga, Spain; ⁶⁰Lab Methabopathies, Centro de Bioquímica y Genética Clínica, H.U. Virgen de la Arrixaca, Murcia, Spain; ⁶¹Unidad de Metabopatías del Hospital Universitario Virgen del Rocío, Seville, Spain; ⁶²National Taiwan University Hospital, Taipei, Taiwan; ⁶³Tennessee Department of Health Laboratory Services, Nashville, Tennessee, USA; ⁶⁴Tanyalcin Medical Lab Selective Newborn Screening and Metabolism Unit, Izmir, Turkey; ⁶⁵Dubai Genetics Centre, Dubai, United Arab Emirates; ⁶⁶Instituto de Seguridad Social, Montevideo, Uruguay; ⁶⁷Washington State Department of Health, Shoreline, Washington, USA; ⁶⁸Wisconsin State Laboratory of Hygiene, University of Wisconsin, Madison, Wisconsin, USA. Correspondence: Piero Rinaldo (rinaldo@mayo.edu)

ratios. Each case is assigned a unique code separate from any other traceable identifier, and no demographic information is collected except the calendar year of birth. Accordingly, this project has been reviewed and approved as a minimum-risk protocol by the Mayo Clinic Institutional Review Board (protocol PR09-001709-01).

The process and criteria used to create a tool are described in the **Supplementary Material** online. Tools can be generated for one or more conditions following a stepwise process that has four major components (**Supplementary Table S1** online): (i) choice of scoring strategy and method to calculate correction factors; (ii) selection of markers; (iii) activation of differentiators, outlier rules, and filters; and (iv) setup of interpretation guidelines. Different scoring strategies are available to elevate scores for conditions that have only a few informative markers (**Supplementary Table S2** online). The correction factors, which can be either condition- or case-specific, are derived from the degree of overlap between the normal population and the disease range of each informative marker in a given condition. The degree of overlap is indeed the foundation of this novel method for interpreting quantitative results in a way that is unique to each condition and therefore not dependent on fixed analyte cutoff values. The selection of markers is based on an objective threshold of clinical significance, which is reached when the median of the disease range of a marker is above the 99 percentile of the normal population (high markers—i.e., abnormal when above the normal range) or below the 1 percentile (low markers).⁶ Differentiators, outlier rules, and filters are added to mitigate the potential impact of true negative cases (cases with completely normal results) to preserve the integrity of the tools and allow differential diagnosis between conditions.

As of 15 December 2011, a total of 90 active tools were accessible on the website, 37 of which are applicable to the differential diagnosis of two or more conditions (**Supplementary Figure S1** online). Their intended use is to generate a score that drives the interpretation and resolution of cases with potentially abnormal MS/MS results. Case profiles can be entered individually (i.e., after the conventional flagging of abnormal results according to cutoff values, **Supplementary Figure S2** online) or as batches containing many profiles (e.g., entire plates/daily runs) uploaded to the website using a health information exchange system.⁸

RESULTS

This multivariate pattern-recognition software is applicable to a broad range of clinical applications. Expanded newborn screening is ideal for a clinical validation study because it involves many markers requiring pattern recognition and profile interpretation. Their complexity is compounded by the rarity of most of the target conditions. At this stage, tools are based on data from neonatal blood spots and are not applicable to different specimen types and to older patients.

Figure 1 shows a partial view of the tool (the data-entry window is not shown; see **Supplementary Material** online for an example of that panel) for argininosuccinic acid lyase deficiency,⁹ a urea-cycle disorder that is included in the recommended

uniform panel.² The top part of the figure is a visual overlay of three elements for each informative analyte (red) and discriminator (gray)—the normal population range, the disease range, and the individual value—all shown after conversion to the multiple of the normal median on a log scale. The screening results of this particular case were not considered informative according to the cutoff value for citrulline applied by the testing laboratory at the time. The bottom part of the figure summarizes the calculated score as follows: (i) the absolute value of the calculated score; (ii) the percentile rank of the score in comparison to all available cases; (iii) the number of available cases with the condition under evaluation; and (iv) a visual display of all scores in comparison to interpretation guidelines. These are built as intervals where the score is considered as either being not informative or indicating that the condition is possible, likely, or very likely. Notably, in this false-negative case the score percentile rank was 29% ($N = 78$) even with the omission from the tool of the unique marker of this condition, argininosuccinic acid. Following this event, the cutoff value of the program was reduced by 25% and this tool is being used on a regular basis. As of 15 December 2011, 110 of the laboratories participating in the collaborative project have implemented a high cutoff value for citrulline and therefore are bound to encounter cases with hypercitrullinemia in disorders besides citrullinemia type I; the most common among these is indeed argininosuccinic acid lyase deficiency. Because 63% of laboratories have a cutoff value above the recommended target range for citrulline (30–40 $\mu\text{mol/liter}$),⁶ they are likely to experience false-negative events like the one shown here. This is not a rare situation. The project database includes 86 cases (0.7% of the total count) that were reported as normal but in which a later diagnosis was based on clinical presentation. This set of cases is limited to those for which all the results required to calculate a score were available, but there are others, some extracted from the literature, with partial sets of data. Excluding conditions in which the poor sensitivity is driven by either a true lack of an informative marker (nonketotic hyperglycinemia) or the historical reliance on an ineffective marker (tyrosinemia type I),⁶ 88% of the remaining cases (59 of 67 patients affected with 23 conditions) generated an informative score when evaluated with the pertinent tool. Six of the eight false-negative cases with uninformative scores have been published.^{10–12} Overall, this anecdotal evidence suggests that, pending a prospective study of the impact of the interpretive tools, at least half of historical false-negative events could perhaps have been avoided if these tools had been available and utilized.

Although sensitivity is of critical importance, the greatest opportunity for performance improvement in newborn screening, especially in a multiplex test environment, is found in the realm of specificity. The false-positive rate limited to testing by tandem mass spectrometry ranges between 5.99% and 0.03% (median: 0.46%) among the 68 sites that have shared their performance metrics on the project website. A significant issue that drives high false-positive rates is the referral to follow-up of newborns with abnormal results due to heterozygosity (carrier status), a situation not uncommon for disorders such as



Figure 1 Partial display of the tool for argininosuccinic acid (ASA) lyase deficiency (two of the three panels; see Supplementary Material online for an example of the data-entry panel). This case was considered not informative on the basis of a cutoff for citrulline set inappropriately high. The top panel is an overlay graph of normal population, disease range, and the values entered to calculate a score. All values are expressed as $\mu\text{mol/l}$ and converted to multiples of the normal median on a log scale. The bottom panel shows the calculated score, the percentile rank comparison to all available scores and the case count along with a graphic display of all available scores for the chosen condition, and a summary of interpretation guidelines.

medium-chain acyl-CoA dehydrogenase deficiency¹³ and very-long-chain acyl-CoA dehydrogenase deficiency.¹⁴ In both conditions, energy depletion due to prolonged labor and delivery may trigger the transient appearance of a biochemical phenotype mimicking affected status. Interpretive tools can facilitate the identification of carriers and consequently reduce the number of cases requiring follow-up. At the same time, use of these tools could prevent at least some of the false-negative events determined by cutoff values set inappropriately, as mentioned

above, but they are not likely to recognize cases with completely uninformative biochemical phenotypes.^{10–12,15} Although several acylcarnitine species could be informative for the evaluation of these conditions,⁶ the most widely used markers are octanoylcarnitine (C8) and tetradecenoylcarnitine (C14:1), respectively. **Figure 2a** shows the distribution of paired C8 and C14:1 concentrations in four groups of cases: two with medium-chain acyl-CoA dehydrogenase deficiency (affected and carriers) and two with very-long-chain acyl-CoA dehydrogenase deficiency

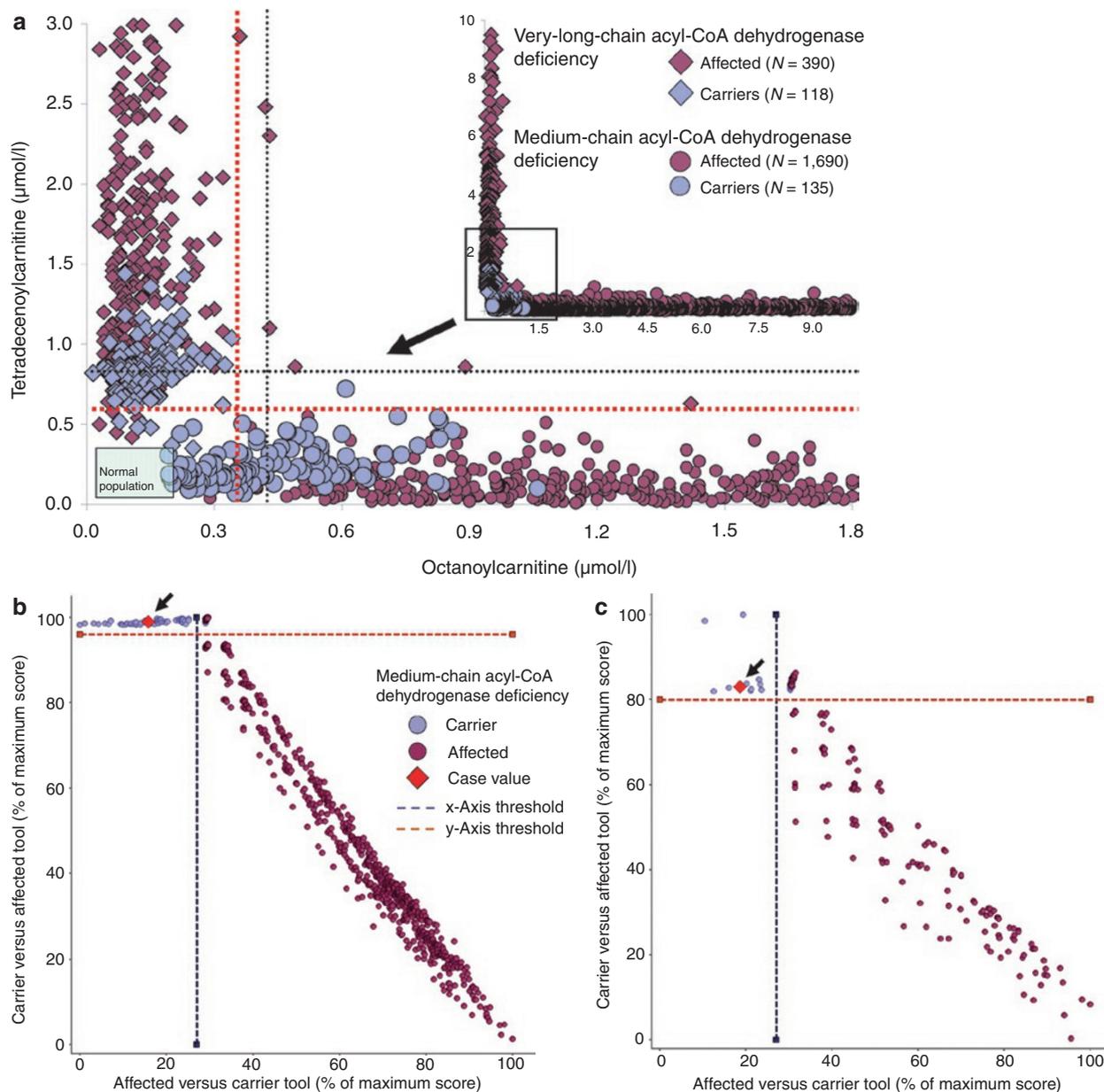


Figure 2 Scatter plots of acylcarnitine results and of condition scores. (a) Scatter plot of C8 and C14:1 in four conditions: medium-chain acyl-CoA dehydrogenase (MCAD) deficiency, heterozygote carriers of MCAD deficiency, very-long-chain acyl-CoA dehydrogenase (VLCAD) deficiency, and heterozygote carriers of VLCAD deficiency. The number of cases included in the figure are shown in the insert in the upper right corner, which shows a wider range of values in affected patients (highest values for C8 and C14:1 are $61.8 \mu\text{mol/l}$ and $13.1 \mu\text{mol/l}$, respectively). The horizontal and vertical red dotted lines correspond to the median cutoff value among all laboratories (C8 $0.35 \mu\text{mol/l}$, $N = 119$; C14:1 $0.60 \mu\text{mol/l}$, $N = 113$). The horizontal and vertical black dotted lines correspond to the median of the ranges in the two carrier groups (C8 $0.44 \mu\text{mol/l}$, $N = 147$; C14:1 $0.84 \mu\text{mol/l}$, $N = 123$). (b) Dual scatter plot comparing the scores of MCAD deficiency (dark circles) and MCAD-deficiency heterozygote carriers (light circles). The dotted lines shown as x-axis and y-axis thresholds define the quadrants of the plot where a combined score (x-axis: $<27\%$, y-axis: $>96\%$; x-axis: $>27\%$, y-axis: $<96\%$) is consistent with carrier status and affected status, respectively. The upper right quadrant defines the small area where a combined score is not informative to discriminate carrier vs. affected (hence to be resolved by biochemical and molecular testing); the lower left quadrant is consistent with normal status. (c) Dual scatter plot comparing the scores of VLCAD deficiency (dark circles) and VLCAD-deficiency heterozygote carriers (light circles). The dotted lines shown as x-axis and y-axis thresholds define the quadrants of the plot where a combined score (x-axis: $<25\%$, y-axis: $>85\%$; x-axis: $>25\%$, y-axis: $<85\%$) is consistent with carrier status and affected status, respectively. Symbols are the same as in panel b. The upper right quadrant defines the area where a combined score is not informative to discriminate carrier vs. affected (to be resolved by biochemical and molecular testing); the lower left quadrant is consistent with normal status.

(affected and carriers). The figure also shows the median values of the two carrier ranges; both values are clearly above the median of all active cutoff values in the collaborative project.

These data illustrate how common it may be to encounter an abnormal result due to heterozygosity, a dilemma that cannot be ignored by increasing the cutoff above the carrier range.¹⁶

The clinical utility of the two conditions tools and, when paired appropriately, of the dual scatter plot is illustrated in panels **b** and **c** of **Figure 2**. They show the scores of two cases, each generated by a tool based on the same markers but designed to recognize the differences between a target condition (affected) and a secondary condition (carrier). A red diamond symbol marks the location of the combined scores of a medium-chain acyl-CoA dehydrogenase deficiency carrier with a concentration of C8 exactly at the median of the carrier range, **Figure 2c** does the same for a very long-chain acyl-CoA dehydrogenase deficiency carrier. Cases with values below the median generate scores that are even more segregated, suggesting that, as in the opposite scenario of false-negative events described previously, in at least half of these cases referred to follow-up the cost of unnecessary tests and a variety of unfavorable outcomes¹⁷ could have been prevented. Furthermore, the application of the same postanalytical process to acylcarnitine profiles generated *in vitro* under controlled circumstances with the fatty-acid probe assay¹⁸ results in a complete separation between the two groups (data not shown). A systematic use of this tool to integrate biochemical and enzymatic results in cases with inconclusive genotyping results has the potential to resolve existing differences of opinion regarding the proper way to follow up

an abnormal newborn screening result.^{19,20} The *in vitro* work is beyond the scope of this report and will be published separately (E.H. Smith, D. Matern, *et al.*, unpublished data).

The impact of this objective, evidence-driven approach to the interpretation of laboratory results could be substantial. As an example, **Figure 3** shows a longitudinal summary of the performance metrics of newborn screening by tandem mass spectrometry in Minnesota over the period 2002–2010. Minnesota has been the first adopter of all quality-improvement tools made available to the participants of the collaborative project since 2005. The first panel shows the number of true-positive cases per year normalized per 100,000 births. Cases of Hmong ethnicity with 2-methylbutyryl-CoA dehydrogenase deficiency²¹ (2–19 new cases per year) were not included in this metric to eliminate the bias of a common disorder in an overrepresented minority. The other two panels show a trend over time of sustained improvement of two performance metrics as described previously,⁵ both greatly exceeding the proposed targets of adequate performance (false-positive rate: 0.30%; positive predictive value: 20%).

DISCUSSION

The primary objective of the Region 4 collaborative project is to promote improvement of laboratory quality of newborn

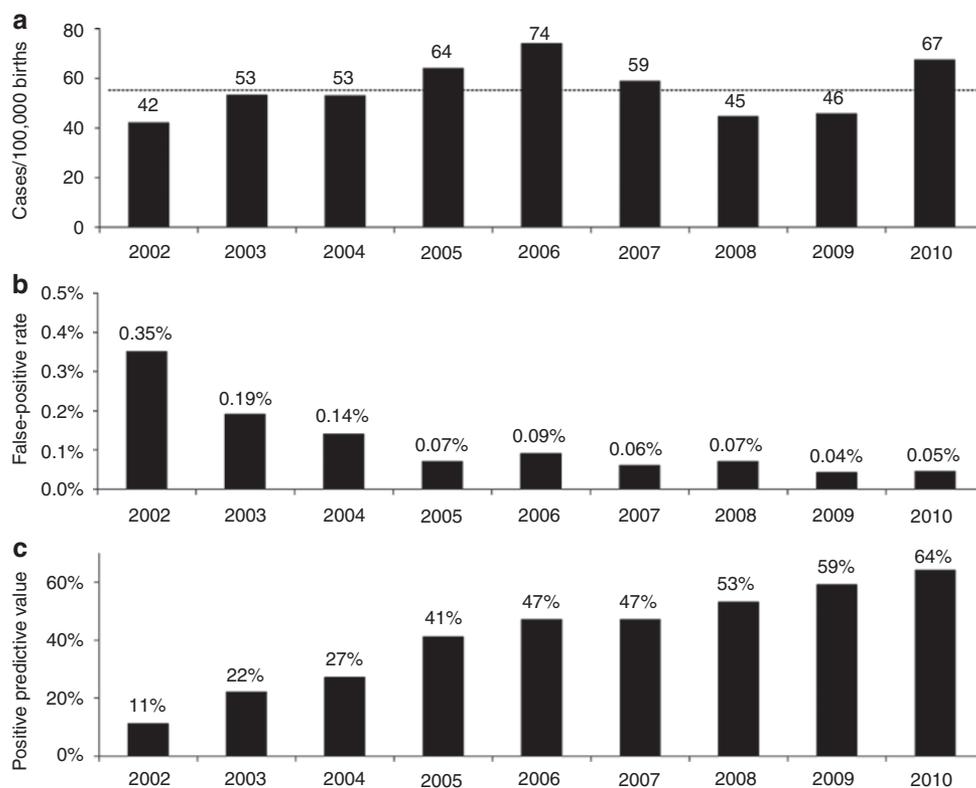


Figure 3 Performance metrics of expanded newborn screening by tandem mass spectrometry in Minnesota, 2002–2010. The birth volume is between 68,000 and 74,000 per year. As result of an ongoing public–private partnership, testing by tandem mass spectrometry was transferred from the Minnesota Department of Health to the Mayo Clinic College of Medicine in June 2004. **(a)** Number of true-positive cases per year normalized to 100,000 births. The dotted line indicates the average (55.8/year/100,000 births). **(b)** Trend of false-positive rate. This metric is expressed as the proportion of positive tests in subjects proven by follow-up evaluation not to have one of the conditions targeted by the Minnesota program.⁷ **(c)** Trend of positive predictive value. This metric is expressed as the probability that a newborn is affected with a condition when restricted to cases with a positive test.⁷ Between June 2004 and December 2010, no false-negative events were brought to the attention of the program with respect to a condition included in the uniform panel.³

screening by tandem mass spectrometry. Dealing with rare conditions of undetermined prevalence, a database of meaningful clinical utility could be produced only through an unprecedented level of cooperation and collaboration on a global scale. The database has led to a new and original type of interpretive tool to achieve reduction in both false-negative events and false-positive outcomes. More traditional statistical methods for separation of cases from noncases, such as likelihood ratio methods and discriminant analysis,²² are not appropriate in the current situation because they assume a multivariate normal distribution of the analyte values in the cases that is not observed. Although many of the detected disorders arise from mutations in a single gene, the variability of the mutations and the extent of the corresponding phenotypic variation are unknown. As such, most of the disease populations are complex mixtures that cannot be modeled with simple parametric distributions. Reliable information for some of the required characteristics, for example, the prevalence of the disease and the complexity of the differential diagnosis needed for a majority of the informative markers, is also lacking. A further disallowing complexity is the number of covariance parameters to be estimated, which vastly exceeds the number of cases of all but the most common of the disorders, making the parametric distribution subject to significant bias.

The lack of traditional analyte cutoff values may seem counterintuitive for reporting quantitative laboratory test results on which binary decisions will be based. However, the basic tenet of this multivariate pattern-recognition software is that an abnormal result is not defined exclusively by a deviation from a statistical definition of normal. The software also evaluates how consistent a result is with the analyte disease range established separately for each condition, an assessment that is novel and more informative than a traditional “one size fits all” cutoff value, and is made possible by a database of true-positive cases of unprecedented size. Another distinctive advantage of the postanalytical tools is the opportunity to calibrate any decision with an element that has not been taken full advantage of so far, which is the degree of overlap between normal population and disease range.

The interpretive tools first became available in January 2009. A conservative estimate of the utilization of the versions based on static spreadsheets is on the order of tens of thousands of downloads; more than 17,000 page views have been recorded since the initial release of the online tools (23 March 2011). The feedback from a diverse spectrum of users, laboratorians, and clinicians has been consistently positive, with indications that these tools are now used in clinical practice on a regular basis and indeed are effective, providing independent verification of the single-site evidence shown in **Figure 3**. A sustained trend of constant improvement is significant because cost–benefit analysis, expense management, and optimization of resource utilization are high priorities in these times of increasing financial constraints, and the public health infrastructure is not exempt from the demand for reducing the cost of health-care services. Future recommendations to expand the uniform newborn

screening panel with the addition of more conditions^{23–25} will raise this pressure even more.

This approach is flexible by design and certainly not limited to amino acids and acylcarnitines. It has already been successfully applied to other multianalyte profiles currently used as either primary or second-tier newborn screening tests, for example, for the interpretation of steroid profiles in congenital adrenal hyperplasia²⁶ and of C₂₀-C₂₆ lysophosphatidylcholine species in X-linked adrenoleukodystrophy and other peroxisomal disorders.²⁷ The availability of more diverse applications is limited only by the gathering of sufficient data of the normal population and of patients affected with the target condition(s).

The software continues to incorporate improvements suggested by users, for example, the ability to customize the pool of percentiles and affected cases relied on to calculate scores. Users have the option to display scores based on subgroups of cases, either their own cases, those belonging to a specific country or, in the future, contributed by laboratories having the closest participant profile in terms of analyte percentiles in the normal population. Additional functions scheduled to be released in the near future are an “all conditions” tool (an unrestricted evaluation of full amino acid and acylcarnitine profiles to suggest any possible diagnosis) and interfaces to download entire batches of raw data from existing commercial software. Additional applications unrelated to newborn screening will become routinely available to span a broad spectrum of either clinical or research endeavors. This evidence-based approach could add substantial value to patient care by providing a comprehensive interpretation of complex laboratory profiles driven by cumulative/multisite evidence and by objective peer comparison.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

Additional public health and private programs have contributed data to the collaborative project, but no individual affiliated with these sites met the authorship criteria for this report. US contributors are from Alabama, Arizona, Arkansas, Colorado, Delaware, Hawaii, Illinois, Maine, Mississippi, Montana, Nebraska, New Hampshire, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, Puerto Rico, Rhode Island, South Dakota, Texas, Utah, Vermont, Virginia, West Virginia, and Wyoming. Partial contributions from international programs were from Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, China, Croatia, Czech Republic, Denmark, Germany, Greece, Hungary, India, Ireland, Italy, Japan, Malaysia, Mexico, Norway, Poland, Qatar, Saudi Arabia, South Africa, Spain, Sweden, Switzerland, Turkey, United Arab Emirates, and the United Kingdom.

This work was supported by a grant (U22MC03963) to the Region 4 Genetics Collaborative from the Health Resources and Service Administration of the Maternal and Child Health Bureau Cooperative Agreement; by contracts from the Eunice Kennedy Shriver National Institute of Child Health and Human

Development, National Institutes of Health, Department of Health and Human Services (contract HHSN275201000017C) and the Newborn Screening Translational Research Network (subcontract HHSN275200800001C 01); and by the T. Denny Sanford Professorship fund, Mayo Clinic College of Medicine.

REFERENCES

1. Wilcken B, Wiley V, Hammond J, Carpenter K. Screening newborns for inborn errors of metabolism by tandem mass spectrometry. *N Engl J Med* 2003;348:2304–2312.
2. Watson MS, Mann MY, Lloyd-Puryear MA, Rinaldo P, Howell RR. Newborn screening: Toward a uniform screening panel and system—Executive summary. *Genet Med* 2006;8(suppl):15–115.
3. Koppaka R. Ten great public health achievements—United States 2001–2010. *MMWR* 2011;60:619–623.
4. Sebelius K. Response by the HHS Secretary to the February 25, 2010 and November 22, 2009 letters. http://www.hrsa.gov/heritabledisorderscommittee/correspondence/response5_21_2010.pdf.
5. Rinaldo P, Zafari S, Tortorelli S, Matern D. Making the case for objective performance metrics in newborn screening by tandem mass spectrometry. *Ment Retard Dev Disabil Res Rev* 2006;12:255–261.
6. McHugh DM, Cameron CA, Abdenur JE, et al. Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. *Genet Med* 2011;13:230–254.
7. Howell RR. Quality improvement of newborn screening in real time. *Genet Med* 2011;13:205.
8. Downs SM, van Dyck PC, Rinaldo P, et al. Improving newborn screening laboratory test ordering and result reporting using health information exchange. *J Am Med Inform Assoc* 2010;17:13–18.
9. Sreenath Nagamani SC, Erez A, Lee B. Argininosuccinate lyase deficiency. In GeneReviews (Internet). Pagon RA, Bird TD, Dolan CR, et al. (eds). University of Washington: Seattle, 1993. <http://www.ncbi.nlm.nih.gov/sites/GeneTests/>. Accessed 3 June 2011.
10. Gallagher RC, Cowan TM, Goodman SI, Enns GM. Glutaryl-CoA dehydrogenase deficiency and newborn screening: retrospective analysis of a low excretor provides further evidence that some cases may be missed. *Mol Genet Metab* 2005;86:417–420.
11. Puckett RL, Lorey F, Rinaldo P, et al. Maple syrup urine disease: further evidence that newborn screening may fail to identify variant forms. *Mol Genet Metab* 2010;100:136–142.
12. Sarafoglou K, Matern D, Redlinger-Grosse K, et al. Siblings with mitochondrial acetoacetyl-CoA thiolase deficiency not identified by newborn screening. *Pediatrics* 2011;128:e246–e250.
13. Matern D, Rinaldo P. Medium-chain acyl-CoA dehydrogenase deficiency. In GeneReviews (Internet). Pagon RA, Bird TD, Dolan CR, et al. (eds). University of Washington: Seattle, 1993. <http://www.ncbi.nlm.nih.gov/sites/GeneTests/>. Accessed 3 June 2011.
14. Leslie ND, Tinkle BT, Strauss AW, Shoener K, Zhang K. Very long chain acyl-Coenzyme A dehydrogenase deficiency. In GeneReviews (Internet). Pagon RA, Bird TD, Dolan CR, et al. (eds). University of Washington: Seattle, 1993. <http://www.ncbi.nlm.nih.gov/sites/GeneTests/>. Accessed 3 June 2011.
15. Sahai I, Bailey JC, Eaton RB, Zytkevich T, Harris DJ. A near-miss: very long chain acyl-CoA dehydrogenase deficiency with normal primary markers in the initial well-timed newborn screening specimen. *J Pediatr* 2011;158:172; author reply 172–172; author reply 173.
16. Maier EM, Pongratz J, Muntau AC, et al. Validation of MCADD newborn screening. *Clin Genet* 2009;76:179–187.
17. Waisbren SE, Albers S, Amato S, et al. Effect of expanded newborn screening for biochemical genetic disorders on child outcomes and parental stress. *JAMA* 2003;290:2564–2572.
18. Matern D. Acylcarnitines, including in vitro loading tests. In: Blau N, Duran M, Gibson KM (eds). *Laboratory Guide to the Methods in Biochemical Genetics*. Springer: Berlin, Germany, 2008:171–206.
19. Arnold GL, Van Hove J, Freedenberg D, et al. A Delphi clinical practice protocol for the management of very long chain acyl-CoA dehydrogenase deficiency. *Mol Genet Metab* 2009;96:85–90.
20. Kronn D, Mofidi S, Braverman N, Harris K; Diagnostics Guidelines Work Group. Diagnostic guidelines for newborns who screen positive in newborn screening. *Genet Med* 2010;12(12 suppl):S251–S255.
21. Matern D, He M, Berry SA, et al. Prospective diagnosis of 2-methylbutyryl-CoA dehydrogenase deficiency in the Hmong population by newborn screening using tandem mass spectrometry. *Pediatrics* 2003;112(1 Pt 1):74–78.
22. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*, 3rd edn. Prentice Hall: Englewood Cliffs, NJ, 1992:552–556.
23. Green NS, Rinaldo P, Brower A, et al.; Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children. Committee Report: advancing the current recommended panel of conditions for newborn screening. *Genet Med* 2007;9:792–796.
24. Perrin JM, Knapp AA, Browning MF, et al. An evidence development process for newborn screening. *Genet Med* 2010;12:131–134.
25. Calonge N, Green NS, Rinaldo P, et al.; Advisory Committee on Heritable Disorders in Newborns and Children. Committee report: Method for evaluating conditions nominated for population-based screening of newborns and children. *Genet Med* 2010;12:153–159.
26. Lacey JM, Minutti CZ, Magera MJ, et al. Improved specificity of newborn screening for congenital adrenal hyperplasia by second-tier steroid profiling using tandem mass spectrometry. *Clin Chem* 2004;50:621–625.
27. Hubbard WC, Moser AB, Liu AC, et al. Newborn screening for X-linked adrenoleukodystrophy (X-ALD): validation of a combined liquid chromatography-tandem mass spectrometric (LC-MS/MS) method. *Mol Genet Metab* 2009;97:212–220.