REVIEW

The role of ERNDIM diagnostic proficiency schemes in improving the quality of diagnostic testing for inherited metabolic diseases

Déborah Mathis[1] | Joanne Croft[2] | Petr Chrastina[3] | Brian Fowler[4] | Christine Vianey-Saban[5] | George J. G. Ruijter[6]

[1]University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[2]Department of Clinical Chemistry and Newborn Screening, Sheffield Children's NHS Foundation Trust, Sheffield, UK

[3]Department of Pediatrics and Inherited Metabolic Disorders, General University Hospital in Prague and First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

[4]University Children's Hospital Basel, University Children's Hospitals Zürich, Basel and Zürich, Switzerland

[5]Division of Inborn Errors of Metabolism, Department of Biochemistry and Molecular Biology, CHU de Lyon, Lyon, France

[6]Department of Clinical Genetics, Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, Rotterdam, The Netherlands

Correspondence
George J. G. Ruijter (on behalf of the ERNDIM Scientific Advisory Board), Department of Clinical Genetics Ee2475, Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.
Email: g.ruijter@erasmusmc.nl

Abstract

External quality assurance (EQA) is crucial to monitor and improve the quality of biochemical genetic testing. ERNDIM (www.erndim.org), established in 1994, aims at reliable and standardized procedures for diagnosis, treatment and monitoring of inherited metabolic disease (IMD) by providing EQA schemes and educational activities. Currently, ERNDIM provides 16 different EQA schemes including quantitative schemes for various metabolite groups, and interpretive schemes such as diagnostic proficiency testing (DPT). DPT schemes focus on the ability of laboratories to correctly identify and interpret abnormalities in authentic urine samples across a wide range of IMDs. In the DPT schemes, six samples each year are distributed together with clinical information. Laboratories choose and perform the tests needed to reach a diagnosis. Data were collected on 345 samples, distributed to up to 105 laboratories worldwide. Diagnostic proficiency (the % of total points possible for all participating laboratories within a scheme for analysis and interpretation) ranged widely: amino acid disorders ($n = 20$), range 33%–100%, mean 84%; organic acid disorders ($n = 35$), range 14%–100%, mean 84%; lysosomal storage disorders ($n = 13$), range 20%–97%, mean 73%; purine/pyrimidine disorders ($n = 9$), range 37%–100%, mean 70%; miscellaneous disorders ($n = 8$), range 17%–100%, mean 65%; no IMD, range 65%–95%, mean 85%. When a sample with the same disorder was distributed in a subsequent survey, performance improved in 75 cases with no improvement seen in 32, suggesting overall improvement of performance. ERNDIM diagnostic proficiency testing is a valuable activity which can help to assess laboratory performance, identify methodological/technical challenges, be informative during quality audits and contribute to a better clinical appreciation of diagnostic uncertainty.

1 | INTRODUCTION

Nowadays, more than 1000 inherited metabolic disorders (IMD) have been identified.[1] Biochemical genetics testing to diagnose these disorders is a demanding task, since they display an enormous biochemical heterogeneity with hundreds of different metabolites present at a wide concentration range. External quality assurance (EQA) plays a crucial role in monitoring and improving the quality of laboratory testing,[2] in particular biochemical genetic testing.[3] ERNDIM (European Research Network for the evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism, www.erndim.org), established in 1994, aims at providing consensus between Biochemical Genetics Centres on reliable and standardized procedures for diagnosis, treatment and monitoring of inherited metabolic diseases.[4,5] This is achieved through EQA schemes operated according to accepted norms on a global scale because in a single country, too few laboratories performing those specific tests exist to obtain sufficient data for reliable performance assessment. The mission of ERNDIM for improvement of biochemical genetic testing is also achieved through educational activities, such as meetings, workshops, training and support.[4] In addition, all EQA scheme reports and recommended operating procedures are freely available online.

Currently, ERNDIM provides 16 different EQA schemes including quantitative schemes for various metabolite groups and interpretive schemes, such as qualitative organic acids in urine (QLOU, see References 5 and 6), qualitative acylcarnitines in dried blood spots (ACDB), mucopolysaccharides in urine (UMPS), congenital disorders of glycosylation (CDG) and diagnostic proficiency testing in urine (DPT). All qualitative schemes focus on the capacity of laboratories to correctly identify and interpret abnormalities in authentic patient samples. Whereas QLOU, ACDB, UMPS, and CDG are each restricted to only one type of test, DPT schemes involve various tests for a wide range of IMDs and are a good approximation of overall routine diagnostic practice. Proficiency testing schemes similar to the ERNDIM DPT schemes have been operated in Australasia[7] and the United States.[8]

In DPT schemes, six samples each year are distributed together with clinical information (Table 1). Laboratories choose and perform the tests needed (limited amount of urine) to reach a diagnosis in consideration of the clinical information given. Laboratories should perform a minimum portfolio of tests to participate in DPT schemes: organic acids, amino acids, purines and pyrimidines, mucopolysaccharides, and oligosaccharides. To allow that, the use of "partner" labs is permitted, although only if this also takes place on a regular basis with real patient samples in routine clinical practice.

There are five different centers that organize DPT schemes: Czech Republic (CZ), France (F), Netherlands (NL), Switzerland (CH), and the United Kingdom (UK).

TABLE 1 Key organizational data of ERNDIM diagnostic proficiency testing (DPT) schemes

Centers organizing DPT schemes	Czech Republic (CZ)
	France (F)
	Netherland (NL)
	Switzerland (CH)
	United Kingdom (UK)
Number of labs per scheme (2020)	19–22
Total number of labs participating (2020)	105
Analysis portfolio needed	Organic acids Amino acids Purines and pyrimidines Mucopolysaccharides Oligosaccharides
Total samples per year	Six samples arranged in two circulations of three samples. One sample is common to all five schemes
Scoring system	Two points for analytical performance Two points for interpretation and recommendation for further tests Total of four points per sample
Analytical performance (%)	Percentage of points for analytics obtained across all labs of the scheme
Interpretational performance (%)	Percentage of points for interpretation and advice for further testing obtained across all labs of the scheme
Overall performance (%)	Percentage of points overall (analytics + interpretation) obtained across all labs of the scheme

As real urine samples are distributed to participants, a limited volume is available and about 7–10 ml of heat-treated urine is sent to each participating laboratory. Each year, there is one common sample distributed across all five DPT schemes. The common samples allow exchange and comparison between the different DPT schemes.

Performance is indicated in an annual combined certificate of participation covering all EQA schemes. Laboratories who score less than 15/24 points, the minimum required for satisfactory performance in DPT schemes, receive a performance support letter issued by the responsible Scientific Advisor of the corresponding scheme. During quality audits, laboratories have to show the relevant inspectors these documents and explain what steps have been taken to address any performance issue(s). The aim is for the participating laboratory to investigate why they scored poorly and for ERNDIM to provide assistance.

In this article, we report on the performance and improved proficiency over the last 15 years across all DPT schemes. Data were collected on 345 samples, including 86 disorders/conditions and distributed to up to 105 laboratories worldwide. We first focus on the overall performance, and secondly on the improvement in proficiency in redistributed samples with the same diagnosis to the same DPT scheme. Assessing the proficiency over time is of utmost importance in judging the value of ERNDIM schemes in sustaining the quality of diagnostic services provided by participating laboratories.

2 | METHOD

2.1 | Scheme organization

Until 2013, DPT surveys were provided by scientific advisors based in academic centers. Since 2014 scheme organization is undertaken by scientific advisors in collaboration with CSCQ (Centre Suisse de Contrôle de Qualité; www.cscq.ch). Scheme organization is coordinated by the ERNDIM administrative office (www.erndim.org).

2.2 | Scoring system in DPT schemes

Performance assessment in the DPT scheme is based on analytical and interpretational aspects. A maximum of four points per sample are given, two points each for analytical performance and interpretation. In the interpretation, recommendations for further testing are also considered. Scoring is performed independently by two

assessors. For all six samples, a maximum of 24 points per year can be achieved by a participating laboratory. Satisfactory performance was set at a minimum score of 15 in any year between 2006 and 2020.

Proficiency is defined, in this study, as the overall combined score for analytical and interpretative proficiency for all participating laboratories for each sample within a scheme, calculated as a percentage.

$$\text{Proficiency} = \frac{\text{Total score of all participating labs for a sample}}{\text{Number of participating labs} \times 4} \times 100\%.$$

Occasionally, where overall proficiency for a given sample is poor, the ERNDIM Scientific Advisory Board (SAB) may classify it as an educational sample, implying that the score is not integrated in the annual performance assessment of the participating laboratories. Samples classified as educational are included in the analyses presented here when scores were available, since repeated circulation of such samples may lead to increased proficiency. Proficiency is used to compare the scores laboratories achieved for each sample, across all five schemes.

2.3 | Fifteen years of data

DPT proficiency data were analyzed over the last 15 years for the CZ, F and CH Schemes (2006–2020), for the NL scheme over the last 12 years (2009–2020), and for the UK scheme over the last 9 years (2012–2020). The data comprised 345 samples, of which 15 were common samples. The data included sample ID, final diagnosis and analytical, interpretational and overall proficiency.

Repeated samples with the same diagnosis were available over the five schemes for 115 samples. The improvement on repeat was assessed as difference in proficiency (delta proficiency = proficiency repeated year − proficiency first year).

3 | RESULTS

3.1 | Diagnostic proficiency among different groups of diseases

The DPT samples (345 samples) were arranged in five different disease groups related to the methods needed to reach the diagnosis. Diagnostic proficiency ranged widely in all disease groups: amino acid (AA) disorders ($n = 20$, 91 samples), range 33%–100%; organic acid (OA) disorders ($n = 35$, 116 samples), range 14–100%;

TABLE 2 Collated proficiency data from 2006 to 2020 over five DPT schemes with change of proficiency in redistributed samples in subsequent survey(s)

Disorders detected with amino acid analysis (20)	n	Proficiency (%) Range	Mean	Δ proficiency on repeat CZ	F	NL	CH	UK
Alpha-amino adipic semialdehyde synthase (AASS) deficiency[a]	1	85	85					
Arginase deficiency	2	87–92	90					
Argininosuccinate lyase (ASL) deficiency	13	72–100	88	+15/+16	+27	−1	+20/+12/+25	+5/+7
Branched-chain aminoaciduria (MSUD)[a]	5	83–100	92	−13/−6		+8	+3	
Citrullinaemia type 1[a]	5	94–99	97	+6	+4/+1		+1/+2	
Cystinuria (cystine/dibasic aminoaciduria)	8	89–100	95	−11/−1		+2		+7
Formiminoglutamic (FIGLU) aciduria (educational: n = 2)	3	33–46	37				+13	
Hartnup disease	2	76–93	85					
HHH syndrome (treated with citrulline)[a]	6	50–84	71	−7	+5		−4/+10	+25
Homocystinuria due to CBS deficiency[a]	7	72–97	88	+21	+4		+6	+8
Hypermethioninemia due to methionine S-adenosyltransferase (MAT) deficiency	2	62–76	69					
Hyperprolineamia type 2	1	77	77					
Hypophosphatasia[a] (educational: n = 1)	6	51–100	81	+27			−17	0
Lysinuric protein intolerance (LPI)[a]	5	80–97	89		+1	−9		+10/+15
Non-ketotic hyperglycinaemia (NKH)	2	87–98	92					
Ornithine aminotransferase (OAT) deficiency	5	92–100	94					+10
Ornithine transcarbamylase (OTC) deficiency	3	77–86	81					
Phenylketonuria (PKU)[a]	4	87–100	97		−12/−3		+1	
Prolidase deficiency (iminodipeptiduria)	9	48–86	71	+38	+9/+11			−6
Tyrosinaemia type 1	2	90–98	93					
Total	91	33–100	84					

Disorders detected with organic acid analysis (35)	n	Proficiency (%) Range	Mean	Δ proficiency on repeat CZ	F	NL	CH	UK
Alkaptonuria (homogentisic acid oxidase deficiency)	5	87–100	96	0				
Aminoacylase 1 deficiency (ACY1D)	4	35–59	49					
Aromatic L-aminoacid decarboxylase (AADC) deficiency (educational: n = 1)	2	14–65	40					
Beta-ketothiolase deficiency (ACAT1)	5	95–97	96		−7		+2	
Canavan disease (N-acetylaspartic aciduria)	3	92–99	95	−4				
Combined malonic and methylmalonic aciduria (ACSF3 gene)	3	54–67	61					
Ethylmalonic encephalopathy (ETHE1)	1	98	98					
Fumarase deficiency	2	78–98	88					
Glutaric acidaemia type 1	9	90–100	95	+1	0/0	+5		
Glycerolkinase deficiency/Xp21 contiguous gene deletion	6	80–100	92		+7/+16			
Hawkinsinuria	1	64	64					
HMG-CoA lyase deficiency	3	87–100	96					

(Continues)

TABLE 2 (Continued)

Disorders detected with organic acid analysis (35)	n	Proficiency (%) Range	Mean	Δ proficiency on repeat CZ	F	NL	CH	UK
2-Hydroxyglutaric aciduria	7	91–100	97	+5	−1			
Hyperoxaluria type 2[a]	1	83–95	88					
Hyperoxaluria type 1	4	73–86	79			+13		
Imerslund Grasbeck (vit B12 malabsorption)	1	84	84					
Isovaleric acidaemia	4	97–100	99				0	
Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency	5	74–100	91					
3-Methylcrotonyl-CoA carboxylase (3MCC) deficiency (3-Methylcrotonylglycinuria)	3	99–100	99					
3-Methylglutaconyl-CoA hydratase (3-MGA) deficiency	2	78–79	78					
3-Methylglutaconic aciduria (Barth Syndrome)	4	66–86	73			+20		
2-Methyl-3-hydroxybutyryl-CoA dehydrogenase (MHBD) deficiency	1	44	44					
Methylmalonic aciduria with homocystinuria (Cbl C)	3	72–78	75				−4	
Methylmalonic aciduria isolated (Mutase, Cbl A)	6	92–100	96			−1		
Methylmalonic semialdehyde dehydrogenase deficiency	1	81	81					
Mevalonic aciduria	4	78–100	93	+8				
Multiple acyl-CoA dehydrogenase (MAD) deficiency	5	85–100	92		−8			
NFU1 (iron–sulfur [Fe–S] clusters) deficiency	1	79	79					
5-Oxoprolinase deficiency	2	91–93	92					−2
Propionic aciduria	4	96–100	99	−4				
Short-chain acyl-CoA dehydrogenase (SCAD) deficiency	4	83–95	90		+6			
Short-chain 3-hydroxyacyl-CoA dehydrogenase (SCHAD) deficiency	2	85–96	91		+11			
Short/branched-chain acyl-CoA dehydrogenase (SBCAD) deficiency	1	79	79					
Succinate semialdehyde dehydrogenase (SSADH) deficiency	5	81–99	90	+8			+7	
Very long-chain acyl-CoA dehydrogenase (VLCAD) deficiency	2	56–71	63					
Total	116	14–100	84					

Lysosomal storage disorders (13)	n	Proficiency (%) Range	Mean	Δ proficiency on repeat CZ	F	NL	CH	UK
Alpha-mannosidosis	7	72–92	80	+14/+19		+6		
Beta-mannosidosis	1	51	51					
Aspartylglucosaminuria (educational: $n = 1$)	8	46–92	69	+17	−10	+9		+39
Fucosidosis	2	62–80	71					
GM1 gangliosidosis	7	65–88	75		+9		−7/−1	
Mucopolysaccharidosis type I (MPS I)	7	82–92	88	−3/−10			+6	

TABLE 2 (Continued)

Lysosomal storage disorders (13)	n	Proficiency (%)		Δ proficiency on repeat				
		Range	Mean	CZ	F	NL	CH	UK
Mucopolysaccharidosis type II (MPS II)	8	75–95	87	0			−1/+5/+8	
Mucopolysaccharidosis type III (MPS III)[a]	9	67–97	80	−4/+12	+12	−30	+10	−2
Mucopolysaccharidosis type IVA (MPS IVA)	7	73–85	80		+9/+10			
Mucopolysaccharidosis type VI (MPS VI)	9	60–98	82	−10/0	+4			+38/+30
Mucopolysaccharidosis type VII (MPS VII)	1	80	80					
Salla disease (SD)[a]	1	20–47	32					
Sialidosis due to neuraminidase deficiency[a] (educational: n = 1)	4	55–88	78	0/+8			−23	
Total	71	20–97	73					

Purine and pyrimidine disorders (9)	n	Proficiency (%)		Δ proficiency on repeat				
		Range	Mean	CZ	F	NL	CH	UK
Adenine phosphoribosyltransferase (APRT) deficiency[a] (educational: n = 1)	2	43–71	57				+15	
Adenylosuccinate lyase (ADSL) deficiency	5	37–62	42	−14				
Dihydropyrimidine dehydrogenase (DPD) deficiency[a]	4	91–100	97			+8	+4	
Dihydropyriminidase (DPH) deficiency	1	80	80					
Lesch–Nyhan disease (HPRT deficiency) (educational: n = 1)	4	41–76	60				+13	
MNGIE/thymidine phosphorylase deficiency (educational: n = 2)	9	45–95	73	+11	+28	+32	+33	
Molybdenum cofactor deficiency	3	51–82	70					
Purine nucleoside phosphorylase (PNP) deficiency	1	74	74					
Xanthine oxidase deficiency	1	74	74					
Total	30	37–100	70					

Miscellaneous disorders or conditions (8)	n	Proficiency (%)		Δ proficiency on repeat				
		Range	Mean	CZ	F	NL	CH	UK
Adenine phosphoribosyltransferase def. + MPS IV	1	50	50					
Cerebrotendinous xanthomatosis (CTX) (educational)	1	36	36					
DOPA therapy	1	91	91					
Essential fructosuria	1	17	17					
Ethylene glycol intake	1	95	95					
Galactosemia	2	51–100	75			+49		
GAMT deficiency[a] (educational: n = 2)	2	59–70	64					
Taurinuria (Red Bull intake)	1	88	88					
Total	10	17–100	65					
No evidence of an IMD	23	65–95	85					
Educational samples not scored (FIGLU, beta-mannosidosis, AGAT deficiency, Wilson disease)	4							
Sample total overall	345							

[a]Disorder circulated as common sample in all five DPT schemes.

TABLE 3 Overall proficiency of the common samples in the different DPT schemes

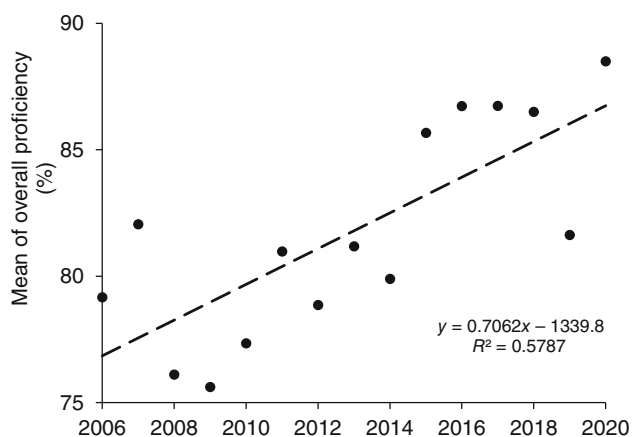|		Overall proficiency (%)							
Year	Diagnosis	CZ	F	NL	CH	UK	Mean	Range	Δ range
2012	Branched chain aminoaciduria (MSUD)	87	83	90	96	83	88	83–96	13
2013	Lysinuric protein intolerance (LPI)	93	88	89	93	82	89	82–93	11
2014	HHH syndrome (treated with citrulline)	62	80	80	70	50	68	50–80	30
2015	Homocystinuria due to CBS deficiency	88	96	84	88	89	89	84–96	12
2016	Hyperoxaluria type 2	95	85	89	85	83	87	83–95	12
2017	Citrullinaemia type 1	99	98	99	98	99	99	98–99	1
2018	Dihydropyrimidine dehydrogenase (DPD) deficiency	100	100	99	97	95	98	95–100	5
2019	Adenine phosphoribosyltransferase (APRT) deficiency	59	48	71	70	43	58	43–71	28
2020	Phenylketonuria (PKU)	98	95	95	99	100	97	95–100	5
2012–2020	Mean proficiency	87	86	88	88	80	86	80–88	

FIGURE 1 The mean of overall proficiency across all schemes over the last 15 years shows a positive trend suggesting improvement in diagnostic testing.

lysosomal storage disorders (LSD) ($n = 13$, 71 samples), range 20%–97%; purine/pyrimidine (PP) disorders ($n = 9$, 30 samples), range 37%–100%; miscellaneous disorders or conditions (misc) ($n = 8$, 10 samples), range 17%–100%; no evidence of an IMD, 23 samples, range 65%–95% (Table 2).

Mean proficiencies were highest for disorders detected with OA and AA analyses (both 84%), followed by LSD (73%), PP disorders (70%), and lastly miscellaneous disorders (65%). Samples with no evidence of an IMD had a mean proficiency of 85% (Table 2).

Very high proficiencies (>97%) were obtained for certain samples in all disease groups, for example cystinuria, alkaptonuria, mucopolysaccharidosis type III (MPS III) and dihydropyrimidine dehydrogenase (DPD) deficiency. In all those samples, very clear biomarker signatures could be detected.

Very low proficiencies (<40%) were obtained for samples with aromatic L-amino acid decarboxylase deficiency (AADC, 14%), fructosuria (17%), Salla disease (SD, 20%), mild homocystinuria due to cystathionine beta-synthase deficiency (CBS, 26%), adenylosuccinate lyase deficiency (ADSL, 28%), formiminoglutamic aciduria (FIGLU, 33%), aminoacylase 1 deficiency (ACY1, 35%) and cerebrotendinous xanthomatosis (CTX, 36%). Samples of guanidinoacetate methyltransferase deficiency (GAMT), L-arginine:glycine amidinotransferase deficiency (AGAT), mitochondrial neurogastrointestinal encephalopathy (MNGIE), adenine phosphoribosyltransferase deficiency (APRT), FIGLU, AADC, CTX, Lesh-Nyhan disease, hypophosphatasia, sialidosis, aspartylglucosaminuria, Wilson disease and beta-mannosidosis, were classed as educational due to poor performance (Table 2).

3.2 | Difference in diagnostic proficiency between DPT scheme centers

We investigated if there was a difference in proficiency between scheme centers. For this we examined the proficiencies of the common samples over the last 9 years (data available for all schemes), to exclude differences resulting from different samples (Table 3). Those data suggest that the scheme organized in the UK performed slightly worse (mean proficiency 80%) than the four other DPT schemes (mean proficiency 86%–88%). For samples with high proficiency (87%–99%), the minimum–maximum range between the different schemes (delta range) was relatively small (1%–13%). On the contrary, if proficiency was relatively low (58%–68%), the range between schemes was higher (28%–30%).
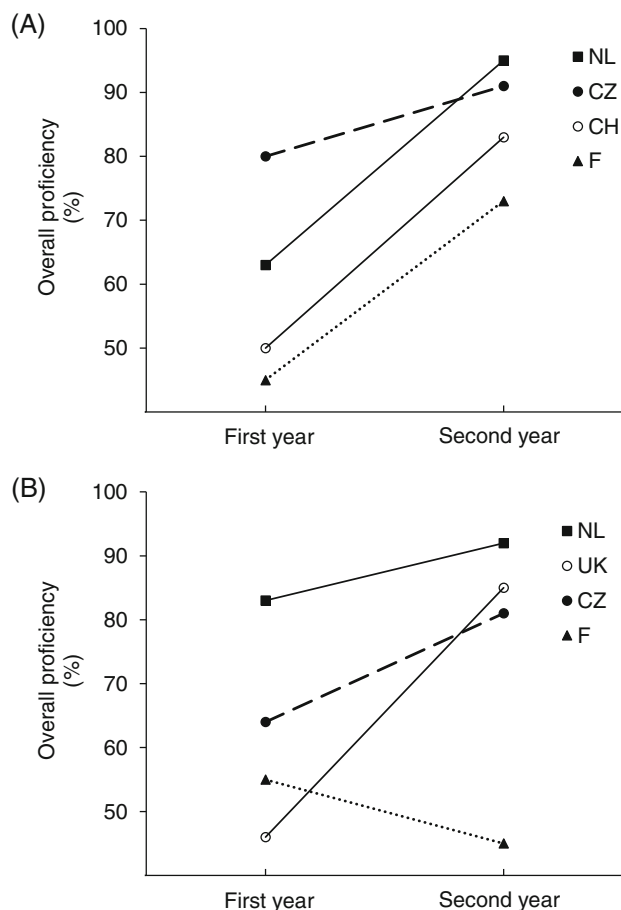
FIGURE 2 Improvement of diagnostic proficiency with repeated sampling in specific samples: (A) MNGIE/thymidine phosphorylase deficiency detected by purine and pyrimidine- and/or organic acid analysis; (B) aspartylglucosaminuria detected by amino acid- and/or oligosaccharide analysis.

3.3 | Improvement of diagnostic proficiency over the years (2006–2020)

The mean of overall proficiency across all schemes over the last 15 years (18–30 samples per year) shows a positive trend (slope = 0.7, R^2 = 0.6, Figure 1) and suggests an overall improvement in proficiency, although this needs to be very carefully interpreted. The same trend was observed for analytical and interpretation proficiency, showing no difference in behavior of those two scoring parameters.

3.4 | Improvement of diagnostic proficiency on repeat sampling

Overall, 115 samples with the same diagnosis have been redistributed in a subsequent survey by the same center.

This takes into account samples with the same diagnosis but not necessarily the exact same sample due to the limited amount of urine available per sample.

When a sample with the same diagnosis has been redistributed in a subsequent survey, performance improved in 75 cases (average increase in score = 13%, range = 1%–49%) (Table 2). Performance remained unchanged in 8 cases and deteriorated in 32 cases (average decrease in score = 7%, range = 1%–30%). We have investigated three cases with decreases in performance in excess of 15%. The deterioration in proficiency of 30% was due to a different MPS III sample being distributed in the later circulation, which did not have a clearly elevated GAG concentration. The deterioration in proficiency of 23% was due to a different Sialidosis sample with a less clear oligosaccharide pattern, which 4 labs of the 11 that performed oligosaccharide analysis classed as GM1. In the first circulation, only 3 labs out of 15 misclassed the samples as GM1. In both surveys, 21 labs were participating in which 15 investigated oligosaccharides in the first survey against only 11 in the second. The reason why less labs performed oligosaccharide analysis is not clear. Finally, the deterioration in proficiency of 17% was due to a different hypophosphatasia sample with a very low concentration of phosphoethanolamine in the second distribution. Two labs that did detect phosphoethanolamine in the second circulation classed it as normal. Nevertheless and overall, the data suggest improvement in performance in redistributed diagnosis in subsequent surveys.

The common sample circulated in 2011 with GAMT deficiency as the diagnosis was classed as educational. In a repeat distribution of a different sample, the DPT-France scheme classed it again as educational, illustrating no clear improvement in the diagnosis of GAMT deficiency. This is probably because the specific biomarkers for GAMT deficiency (guanidinoacetate and creatine) are not included in the standard portfolio for participating in DPT schemes.

Figure 2A illustrates the improvement in proficiency of MNGIE/thymidine phosphorylase deficiency samples (delta proficiency of +11/+28/+32/+33 in the four respective schemes). Participating laboratories were probably more aware of the need to perform purine/pyrimidine analysis given the clinical details after the first distribution. In DPT-France, the first distribution was considered as educational due to the poor overall proficiency of 45%. As another example, aspartylglucosaminuria (delta proficiency +17/−10/+9/+39) in subsequent surveys shows clear improvement especially in the UK scheme (Figure 2B). The sample was classed as educational by the SAB for the UK scheme the first year and performance improved on repeat distribution.

4 | DISCUSSION

4.1 | Overall performance

In this study we report on the performance of laboratories that participated in 2006–2020 in the Diagnostic Proficiency Testing (DPT) EQA schemes provided by ERNDIM. DPT schemes assess diagnostic proficiency related to inherited metabolic disorders (IMD), including test selection, analysis, interpretation, and advice for further testing. The performance of all participating laboratories in the different DPT schemes ranged widely from below 40% for challenging samples up to 100% for more straightforward ones. Moreover, broad variation in proficiency among participating laboratories was observed for diagnoses with relatively low proficiency in contrast to constant high proficiency in samples with more common diagnoses. Even taking into account that a definitive diagnosis is not always possible with urine alone, the performance for some samples was particularly low. Various factors might affect diagnostic proficiency: these can relate to laboratory organization not specific to IMD testing, that is, availability of appropriate methodology and the presence of a quality management system with validation of newly introduced tests, robust laboratory testing processes, and trained personnel. On the other hand, a crucial factor that is particularly important to IMD's and diagnosing rare disorders is related to experience, reflecting numbers and range of specimens received, as well as years of staff involvement. Low proficiency could thus be explained by unfamiliarity of laboratories with more exotic diagnoses or inexperience with samples with less clear metabolite signatures. All of the above factors may contribute to explain the fact that, over those years, 17 samples have not been included in performance assessment, but were judged to be educational due to very poor performance. An additional reason for low profiency may be that the tests to detect the primary biomarkers of these diseases are not included in the standard "portfolio" required to participate in DPT schemes. This may have affected 6 of the 17 educational samples (fructosuria, CTX, GAMT deficiency, AGAT deficiency, and Wilson disease).

The inclusion of disorders of uncertain clinical significance (e.g., FIGLU, aminoacylase 1 deficiency) and disorders where urine testing is not the usual mode of diagnosis (e.g., galactosemia, VLCAD deficiency, CTX, Wilson disease) in the DPT schemes may also contribute to relatively poor proficiency. Laboratories may have actively decided not to target the relevant markers in the former situation. In the latter situation, labs may recommend alternative samples/tests in routine practice and consequently not expend much effort on developing expertise in the analysis and interpretation of urine samples when a test with better clinical and practical utility is locally available.

For distributions of samples with no evidence of an IMD, proficiency ranged between 65% and 95% with a mean of 85, which is not especially high. This may be due to over-interpretation of minor metabolite changes due to, for example, diet, medication or pre-analytical degradation in urine samples. The trend to over-interpret normal samples in EQA setting has already been reported by Peters et al. in 2008 and 2016.[5,6]

In comparison to other interpretative ERNDIM schemes, DPT is surely the most complex scheme, because the participating laboratories must select and perform various tests, including interpretation of (minor) metabolite variations that may be due to non-genetic conditions or pre-analytical changes of the urine sample. However, DPT schemes reflect the real diagnostic situation most adequately and thus, low performance obtained for some diagnoses remains problematic.

Clinicians must be aware that laboratory testing for IMDs can be difficult and results need to be viewed with appropriate caution. Critical tests may not be performed or metabolite abnormalities could have been missed. Repeating some laboratory tests on fresh samples taken at different time points should be more often taken into consideration. This has also been reported by Peters et al. in 2016.[5] Furthermore and in real settings, plasma samples are frequently available and test results from plasma and urine combined are often more informative than urine samples alone.

4.2 | Improvement of performance over the years

The overall performance seems to show a positive trend, although this needs to be carefully interpreted due to broad variation of the DPT samples in each year. In 2008, Fowler et al. reported a clear decrease in the number of poor performers in the CZ and F scheme over the 2002–2007 period.[4] It may be that a larger increase in performance was achieved during the early years of DPT schemes and that the increase in performance slowed down after that.

Here, it should be noted that results for a redistributed sample can only be compared within a single DPT scheme center, as participants remain relatively stable within a particular DPT scheme. Thus, repetition of a diagnosis in a single DPT scheme will best show improvement in proficiency as a result of the learning process in individual laboratories. This is illustrated in Figure 2 with the improvement in proficiency of MNGIE/thymidine phosphorylase deficiency and aspartylglucosaminuria.

A limitation in the interpretation of altered proficiency in repeated samples of the same diagnosis is that the sample may not be the same but from a different patient or from the same patient, but sampled at a different time. Therefore, the signature metabolites might be present in different amounts, making analysis and interpretation more straightforward or less. For example, the repeat sample of MPS III showed a less clear GAG increase and led to decrease in proficiency. Alteration in proficiencies could also be due to methodologies that may have changed, or tests not being performed. The latter was observed with sialidosis, for which less labs investigated oligosaccharides during the second circulation.

How to increase the improvement in proficiency is under discussion by the SAB to address issues in failure to detect very clear metabolite patterns and in interpreting the results. Poor performance mostly arises when less clear metabolite signatures are present in a sample, or in some cases when the test required to detect the specific biomarkers is not part of the standard portfolio of DPT scheme. Therefore, a possibility for proficiency improvement would be to just select samples with clear metabolite signatures that are routinely detected in urine using the standard portfolio of tests required for DPT schemes. However, the utility of circulating not only trivial samples, but also challenging ones, is to start off fruitful discussions of critical aspects about diagnosis paths and analytics, for example methodological issues. There is clearly room for improvement of performance which may be achieved by method improvements and harmonization, use of certified reagents and standards and shared experience with unusual disorders, activities to which ERNDIM is firmly committed. Publishing summarized protocols, both on the ERNDIM website and the literature, including diagnostic clues and pitfalls could be of great help for the labs. Scheme annual reports are a valuable resource and are available online, but there is no guarantee that they are read by the participants. Workshops on DPT and other schemes are well attended and surely contribute to the education and improvement of biochemical genetic laboratory testing. In addition to EQA schemes, ERNDIM also provides control materials for internal QC and an educational kit containing positive oligosaccharidosis samples. As from 2014, the ERNDIM DPT scheme also issues critical errors (CE), and this designation may contribute to improvement in proficiency. CE are defined as an error deemed unacceptable to the majority of laboratories and which would have a serious adverse effect on patient management. For example, failure to perform a relevant test, missing a diagnosis when proficiency for that sample is >95%, or making a misleading conclusion could lead to CE. What

constitutes a CE for each scheme is discussed on a case-by-case basis by the ERNDIM SAB each year. Laboratories who otherwise obtained an acceptable annual score but who have made a CE automatically receive a performance support letter.

Nevertheless, the observed overall trend for improved performance in repeat distributions is encouraging. ERNDIM DPT surveys may have contributed to improvement in diagnostic biochemical genetic testing by providing a wide variety of IMDs, some of which laboratories have not yet encountered in their patient populations. This will enable them to learn the identity and characteristics of previously unidentified metabolites, to enhance their knowledge and experience regarding rare disorders and to improve their diagnostic skills.

5 | CONCLUSION

This work contributes importantly to the evaluation of the current state in diagnosis of a wide range of inherited metabolic disorders within laboratories worldwide. Further, we show an overall trend in improvement of proficiency in ERNDIM DPT schemes over the years, although poor performance for some samples remains. ERNDIM diagnostic proficiency testing is a valuable activity, which may contribute to the improvement of diagnostic biochemical genetic testing worldwide. In addition, it helps to assess laboratory performance, can be informative during quality audits, identifies methodological and technical challenges, and contributes to a clearer clinical appreciation of diagnostic uncertainty.

AUTHOR CONTRIBUTIONS
Déborah Mathis: Scientific advisor (SA) of ERNDIM DPT Switzerland scheme, provided data, computed data and drafted the manuscript with tables and figures. Joanne Croft: SA of ERNDIM DPT United Kingdom scheme, provided data, performed initial review and organization of data, critically reviewed the manuscript. Petr Chrastina: SA of ERNDIM DPT Czech Republic scheme, provided data, critically reviewed the manuscript. Brian Fowler: Former SA of ERNDIM DPT Switzerland scheme, provided data, performed initial review and organization of data, critically reviewed the manuscript. Christine Vianey-Saban: SA of ERNDIM DPT France scheme, provided data, critically reviewed the manuscript. George J. G. Ruijter: SA of ERNDIM DPT Netherland scheme, provided data, critically reviewed the manuscript (correspondence).

no conflict of interest. Christine Vianey-Saban has served in a Scientific Working Group for Ultragenyx.

DATA AVAILABILITY STATEMENT

The data used in this study are available upon reasonable request from the corresponding author.

ETHICS STATEMENT

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. All patient urine samples were pseudonymized and were obtained in accordance with national legislation and institutional guidelines of the different university hospitals.

REFERENCES

1. Ferreira CR, Rahman S, Keller M, Zschocke J. An international classification of inherited metabolic disorders (ICIMD). *J Inherit Metab Dis*. 2021;44:164-177.
2. Sciacovelli L, Secchiero S, Zardo L, Zaninotto M, Plebani M. External quality assessment: an effective tool for clinical governance in laboratory medicine. *Clin Chem Lab Med*. 2006;44(6): 740-749.
3. Hoffman GF. Selective screening for inborn errors of metabolism—past, present and future. *Eur J Pediatr*. 1994;153:2-8.
4. Fowler B, Burlina A, Kozich V, Vianey-Saban C. Quality of analytical performance in inherited metabolic disorders: the role of ERNDIM. *J Inherit Metab Dis*. 2008;31:680-689.
5. Peters V, Bonham JR, Hoffmann GF, Scott C, Langhans CD. Qualitative urinary organic acid analysis: 10 years of quality assurance. *J Inherit Metab Dis*. 2016;39:683-687.
6. Peters V, Garbade SF, Langhans CD, et al. Qualitative urinary organic acid analysis: methodological approaches and performance. *J Inherit Metab Dis*. 2008;31:690-696.
7. Tan IK, Gajra B, Lim MSF. External proficiency testing programmes in laboratory diagnoses of inherited metabolic disorders. *Ann Acad Med Singapore*. 2006;35:688-693.
8. Oglesbee D, Cowan TM, Pasquali M, et al. CAP/ACMG proficiency testing for biochemical genetics laboratories: a summary of performance. *Genet Med*. 2018;20:83-90.