Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

Analytical Challenges, Standardization Strategies, and Future Directions of LC-MS/MS in Neonatal Screening: A Systematic Review

[1]*Binata Shrestha, [2]Sujan Koju, [3]Ekata Shrestha
[1]Gnosis- Mountain View Medical Laboratory(MVML), Department of Toxicology, USA.
[2]Medical Laboratory Technologist, Dhulikhel Hospital, Nepal.
[3]Medical Microbiology, Dhulikhel Hospital, Nepal.

Abstract

The use of liquid chromatography/tandem mass spectrometry has been a key development in expanded newborn screening, as it enables the simultaneous determination of many metabolites from a single dried blood spot collected shortly after birth, and enables both first tier and confirmatory testing. The method is highly analytical specific, but there are still issues to be resolved regarding the harmonization of results, thresholds and methods between laboratories. The aim of this systematic review was to find out the analytical problems, to assess the standardization processes and to sum up the future directions mentioned in the primary literature. Twenty-four studies were identified and included in the systematic search of four electronic databases, mostly from single centers in high-income countries, and appraised. Three interrelated areas of analytical challenge were identified: preanalytical effects of the dried blood specimen such as effect of the hematocrit, biological and exogenous contamination, and preparation of the quality control materials; analytical interferences, primarily isobaric and isomeric species that must be chromatographically separated, and matrix effects and the compromises between derivatized and underivatized methods; and quantitative difficulties associated with population-dependent thresholds and the lack of commutable reference materials. There were four levels of standardization: external quality assessment and proficiency testing, consensus guidelines and accreditation, post-analytical harmonization of covariate-adjusted interpretation of results and analyte-ratio algorithms, and metrological traceability, with the latter two being the most immediate sources of the reductions in false-positive results. The future directions that were reported were focused on high resolution and untargeted profiling, proteomic multiplexing, integration with genomic sequencing, and machine learning. The platform analytics was therefore determined to be well-developed, while harmonisation between laboratories remained the main constraint and the development of commutable reference materials and the addition of underrepresented populations were determined to be priorities for future research.

Keywords: Newborn screening; Liquid chromatography–tandem mass spectrometry; Dried blood spot; Inborn errors of metabolism; Second-tier testing; Analytical harmonization; Cutoff derivation; Proficiency testing.

1. Introduction

Newborn screening is considered one of the most cost effective secondary prevention programs in modern medicine because biochemical screening in the newborn period allows treatment to be started before irreversible neurological damage or death occurs, within the pre-symptomatic window (Guthrie & Susi, 1963). The use of a bacterial inhibition assay to capillary blood dried on filter paper was applied to population-based screening, a sample type that is still used today in screening programs (Guthrie & Susi, 1963). This one disorder origin has evolved into a multifaceted public health system which can identify over fifty conditions (Therrell et al., 2015).

The scope of screening was expanded most decisively when tandem mass spectrometry was introduced into the laboratory workflow. It was shown that acylcarnitine profiles could be obtained from blood spots at physiological concentrations, suggesting that several fatty-acid oxidation and organic-acid disorders could be detected at the same time (Millington et al., 1990). Shortly after, the platform was used to directly quantify phenylalanine and tyrosine in neonatal blood spots, confirming the platform's diagnostic potential (Chace et al., 1993). Currently, amino acids and acylcarnitines are analyzed as a single sample from a single

Binata Shrestha
Gnosis- Mountain View Medical Laboratory(MVML), Department of Toxicology, USA.
Email: binata123@gmail.com

Received: 26-Jun-2026 Revised: 22-Jul-2026 Accepted: 3-Aug-2026

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

dried-blood-spot punch (Ombrone et al., 2016), which is referred to as one specimen, many disorders.

Two analytical configurations are used in practice. First tier screening is carried out mainly by flow-injection-analysis tandem mass spectrometry, where no separation step is performed before the detection step, while the addition of a separation step, such as liquid chromatography, coupled to tandem mass spectrometry (LC-MS/MS) increases the specificity (Ombrone et al., 2016). Therefore, LC-MS/MS has been widely used for second tier testing, where the measurement of methylmalonic acid, methylcitrate, homocysteine and succinylacetone are measured, and for newer targets like lysosomal storage disorders and X-linked adrenoleukodystrophy (la Marca et al., 2008).

Although LC-MS/MS has these benefits, there are some limitations to the analytical reliability of this approach in this context. Pre-analytical variability is mainly due to the dried blood spot matrix, in which the hematocrit determines the spreading of the blood and the amount sampled by a fixed punch, which will result in a bias in quantification (De Kesel et al., 2013). The choice of derivatized or underivatized protocols, as well as the use of stable-isotope-labeled internal standards, and the presence of isobaric interferences and matrix effects, all affect the performance during analysis (Ombrone et al., 2016).

One of the basic problems is the lack of standardization. While analytical specificity is high, there is still a lack of harmonisation of measured results, cut-off values and methods between laboratories, and there are few commutable certified reference materials, which means that false-positive and false-negative rates vary significantly between programs (Pickens et al., 2020). Responses have been coordinated, including the distribution of the reference materials and proficiency-testing schemes through the Newborn Screening Quality Assurance Program (NSQAP) (De Jesús et al., 2015) and the post-analytical, covariate-adjusted interpretive tools, which were developed from multi-site data (McHugh et al., 2011); however, their use is not uniform.

The relevant evidence is scattered throughout the analytical-chemistry, laboratory-medicine and public-health literature, and has not yet been brought together and critically evaluated in a single reporting system. A systematic review is therefore warranted, so that the analytical challenges reported can be mapped, the strategies for standardisation that have been assessed can be evaluated and the future directions that are proposed in the primary studies can be identified. This synthesis creates an evidence base that is appraised and integrated and provides information to laboratories, screening authorities and policymakers.

The present study is centered on the following objectives:

- To systematically identify and categorize the analytical challenges reported for LC-MS/MS in neonatal screening across the pre-analytical, analytical, and post-analytical phases.
- To evaluate the standardization and harmonization strategies that have been assessed in the primary literature, together with their reported effectiveness.
- To synthesize the future directions proposed in primary studies, spanning instrumentation, panel expansion, genomic integration, and computational interpretation.

Accordingly, the study is guided by the following research questions:

RQ1. What analytical challenges are reported for LC-MS/MS in neonatal screening, and how do these challenges affect diagnostic performance?

RQ2. Which standardization and harmonization strategies have been evaluated, and with what demonstrated effect on inter-laboratory comparability and on false-positive and false-negative rates?

RQ3. What future directions are identified in the primary studies for improving LC-MS/MS-based neonatal screening?

2. Materials and Methods

2.1 Eligibility criteria

Eligibility criteria were pre-defined and organized in a Population, Exposure, Comparator and Outcome (PECO) structure, as per the PRISMA 2020 statement (Page et al., 2021). The population consisted of neonates undergoing screening by dried blood spot or other microsample, and the exposure consisted of using liquid chromatography–tandem mass spectrometry (LC-MS/MS) as a first-tier, second-tier, or confirmatory procedure, in which the tier of screening was interpreted according to consensus guidance (Hoffman, 2010).

Comparisons to other methods reported as eligible comparators included flow-injection-analysis tandem mass spectrometry, immunoassay, fluorometric assays, and other reference methods. Analytical-performance parameters, harmonization and standardization parameters, false-positive and false-negative rates, and method-comparison estimates were the outcomes of interest. Studies were eligible if they were method-development, analytical-

validation, method-comparison, external-quality-assessment, proficiency-testing, or program-evaluation reports; quantitative or qualitative data could be extracted. Only English publications were retained and the search was limited to those published since the introduction of tandem mass spectrometry into neonatal screening. Studies were excluded if there was no neonatal matrix available for analysis, if the study was a purely diagnostic or clinical cohort, if it was a narrative review, editorial, commentary, or conference abstract, and if primary data were not available. The inclusion and exclusion criteria are summarized in Table 1.

Table 1. Inclusion and exclusion criteria (adapted PECO framework)

Domain	Inclusion criteria	Exclusion criteria
Population / specimen	Neonates screened using dried blood spots or equivalent microsamples	Non-neonatal populations; non-neonatal matrices
Exposure (index technology)	LC-MS/MS used as a first-tier, second-tier, or confirmatory method	Techniques not involving LC-MS/MS; unspecified platform
Comparator	FIA-MS/MS, immunoassay, fluorometric, or other reference methods, where reported	Not required; absence of a comparator did not exclude a study
Outcomes	Analytical performance, harmonization/standardization metrics, false-positive and false-negative rates, method-comparison estimates	No extractable analytical, standardization, or performance outcome
Study design	Method-development, validation, method-comparison, EQA, proficiency-testing, program-evaluation reports	Narrative reviews, editorials, commentaries, abstracts without extractable data
Language and period	English; from the introduction of MS/MS in neonatal screening onward	Non-English records; records predating the defined period

2.2 Information sources and search strategy

PubMed/MEDLINE, Scopus, Europe PMC, and the Cochrane Library are the four selected databases. The search was planned and reported using the PRISMA 2020 statement and literature search reporting extension (Page et al., 2021; Rethlefsen et al., 2021). In addition to the primary databases, guideline repositories and grey literature were consulted, such as newborn-screening quality-assurance documentation, to ensure that material not indexed in the primary databases was captured.

Three conceptual blocks were used to create each search string, which were then linked together by the Boolean operator AND: the analytical technology, the screening context and specimen, and the analytical-performance or standardization terms. Controlled vocabulary, such as Medical Subject Headings, was used in each block in combination with the free-text synonyms using the operator OR, and truncation was used to capture lexical variants. The complete, database-specific strategies are presented in Table 2.

Searches were performed on a specific date only, indicated along with the results, and no automated language or publication filters were used at the database level other than those inherent in the eligibility criteria. All retrieved records were exported to a reference-management software for which duplicates were eliminated prior to screening. To improve the sensitivity of retrieval, the reference lists of the included studies and of pertinent reviews were manually screened, so that additional eligible reports could be identified and added to the study pool (Page et al., 2021).

2.3 Data extraction

A standardized form was used to extract data, which was tested on a sample of the studies included and modified before it was used in the full study. The extracted fields included the screening program, the target disorders and analytes, the instrumental configuration, the chromatographic column chemistry, the use of a derivatized or underivatized protocol, the use of a stable-isotope-labeled internal standard(s), and the approach for the calibration and the derivation of the cutoff.

The analytical-performance parameters were noted based on the characteristics established for liquid chromatography–mass spectrometry methods, such as the limits of detection and quantification, linearity, precision, recovery, ion suppression, carry-over, and analyte stability

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

Table 2. Database-specific search strategies

Database	Search Strategy
PubMed/MEDLINE	(("Mass Spectrometry, Tandem"[Mesh] OR LC-MS/MS OR "liquid chromatography tandem mass spectrometry")) AND ("Newborn Screening"[Mesh] OR "newborn screening" OR "neonatal screening" OR "dried blood spot" OR DBS) AND ("method validation" OR validation OR standardization OR harmonization OR "quality assurance" OR calibration OR "reference materials" OR "analytical performance" OR "matrix effects" OR "ion suppression")
Scopus*	TITLE-ABS-KEY(("LC-MS/MS" OR "liquid chromatography tandem mass spectrometry" OR "tandem mass spectrometry" OR MS/MS) AND ("newborn screening" OR "neonatal screening" OR "dried blood spot" OR DBS) AND (validation OR "method validation" OR standardization OR harmonization OR "quality assurance" OR calibration OR "reference materials" OR "analytical performance" OR "matrix effect*" OR "ion suppression"))
Europe PMC	("LC-MS/MS" OR "liquid chromatography tandem mass spectrometry") AND ("newborn screening" OR "neonatal screening" OR "dried blood spot") AND (validation OR standardization OR harmonization OR "quality assurance" OR calibration OR "analytical performance" OR "method validation" OR "matrix effect" OR "ion suppression")
Cochrane Library	("newborn screening" OR "neonatal screening" OR "dried blood spot" OR DBS) AND ("mass spectrometry" OR "tandem mass spectrometry" OR LC-MS/MS OR MS/MS OR "liquid chromatography")

(Clinical & Institute, 2014). The matrix effects were removed as reported, following the approach taken for the assessment of matrix effects in quantitative bioanalytical methods (Matuszewski et al., 2003). Extraction was performed independently by two reviewers in order to minimize transcription error.

2.5 Data synthesis

Owing to the anticipated methodological and clinical heterogeneity, a narrative and thematic synthesis was undertaken and reported in accordance with the Synthesis Without Meta-analysis (SWiM) guideline (Campbell et al., 2020). The results were summarized in structured summary tables around the three review questions. Quantitative pooling was reserved for instances in which a sufficient number of methodologically comparable studies reported concordant outcome measures; where pooling was not appropriate, the heterogeneity and the rationale for a descriptive approach were reported transparently (Page et al., 2021).

3. Results

3.1 Study selection

The study identification and selection process is summarized in the PRISMA 2020 flow diagram (Figure 1) which was created to follow current reporting guidance (Page et al., 2021). The systematic search of the four electronic databases yielded a total of 1,942 records (Europe PMC: n = 990, Scopus: n = 540, PubMed/MEDLINE: n = 310, Cochrane Library: n = 102) and no records were found from trial registers. Prior to screening, 1,364 records were eliminated, including 890 that were duplicate records, 301 that were ineligible due to automation tools, and 173 that were removed for other reasons.

After removal of these records, 578 unique records were then examined at the title and abstract level for eligibility based on the eligibility criteria that were developed. Of these, 295 were found to be ineligible for inclusion in the study: most frequently due to an ineligible population, a non-neonatal matrix, or a lack of an LC-MS/MS component. The full texts of the remaining 283 reports were then requested for retrieval and 167 of these were not retrieved.

Of the 116 reports successfully retrieved, full assessment of the reports to the eligibility criteria was conducted. At this stage, 92 reports were excluded, the reasons were recorded and categorised according to the eligibility framework, which included an ineligible population or specimen matrix, absence of LC-MS/MS as the index method and absence of an extractable analytical, standardization, or performance outcome, as well as records that were reviews, editorials, or conference abstracts.

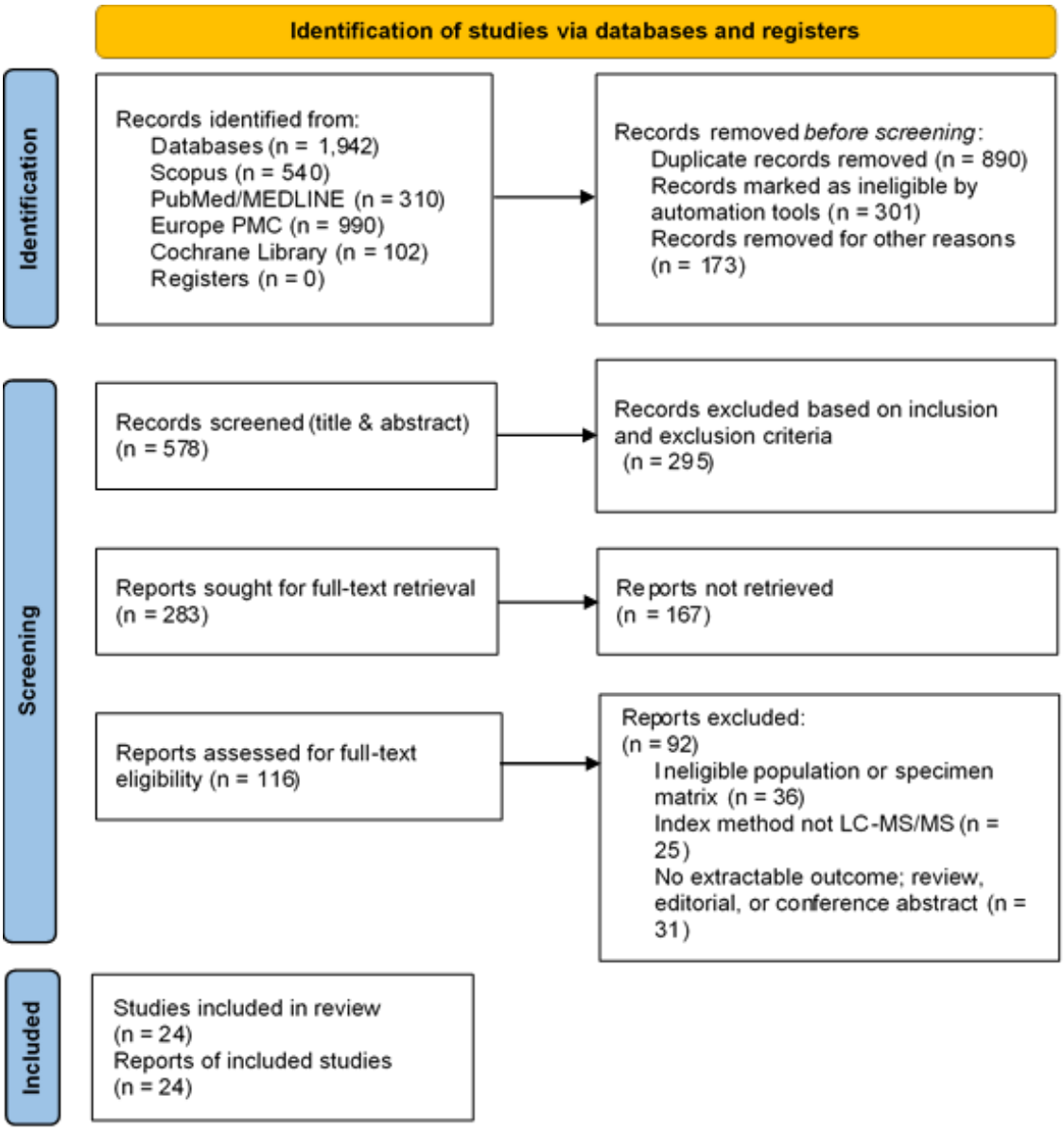

Figure 1. PRISMA framework diagram

A total of 24 studies fulfilled all eligibility criteria and were included in qualitative synthesis, each of which was represented by one primary report (Figure 1). The included studies were published between 2008 and 2026 and originated from a geographically diverse, although predominantly high-income, set of screening programs. One record was retained as a borderline inclusion, as neonatal dried blood spots and LC-MS/MS were used, but the analytical target was actually a research application (not a screening-panel application); one additional study was for dried urine as an equivalent neonatal microsample.

3.2 Characteristics of included studies

The principal characteristics of the 24 included studies are summarized in Table 3. In addition to the analytical method-development and validation reports (Hong et al., 2018; Hubbard et al., 2009; Kilgore et al., 2023; Klippel et al., 2025; Oguni et al., 2020; Posern et al., 2024; Stapleton et al., 2020; Wang et al., 2013), the corpus was complemented by program evaluations (Engström et al., 2026; Fecarotta et al., 2025; la Marca et al., 2008), an external-quality-assessment material-development study (Courtney et al., 2026), a quality-control method comparison (Grecsó et al., 2021), reference-value studies

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

(Abdulridha et al., 2024; Saeed et al., 2024), and method comparisons incorporating machine learning (Mak et al., 2023; Xie et al., 2025).

The studies were geographically located in the United States (n = 7), Italy (n = 3), Japan (n = 2), Canada (n = 2), Iran (n = 2) and also from one study in Spain, Sweden, Hungary, Germany, Thailand, and Denmark. The two Iranian reference-value studies were in middle income countries while the other programs were in high income countries. LC-MS/MS was used as a first-tier, second-tier or confirmatory method across the corpus and several studies integrated the platform with other technologies such as high-resolution metabolomics, next-generation or genome sequencing, and MALDI-TOF peptide profiling (Courraud et al., 2021; Fecarotta et al., 2025; Phoungphosop et al., 2026; Xie et al., 2025).

The disorders covered included second tier metabolic panels, mucopolysaccharidoses and other lysosomal disorders, X-linked adrenoleukodystrophy, congenital adrenal hyperplasia and emerging disorders like Wilson disease, Canavan disease and congenital hypothyroidism. In most studies, the specimen was dried blood spot, in one study it was dried urine (Menkovic et al., 2019) and in another it was a mixed plasma and dried-blood-spot design (Sidorina et al., 2026). Single-center or single-program designs were observed, and there were very few truly reproducible studies that involved multiple sites.

Table 3. Characteristics of the included studies

#	Study (Year)	Country / program	Target analytes → disorders	Tier	Design / method type
1	Courraud et al. (2021)	Denmark (iPSYCH biobank)	~4,360 metabolite features → ASD (research)	Research on NBS DBS	Nested case–control; untargeted metabolomics (HRMS)
2	Courtney et al. (2026)	USA (CDC NSQAP)	Recombinant enzyme substrates → MPS I/II, Gaucher, Fabry, Krabbe, Pompe, ASMD	First-tier PT/QA	EQA / PT-material development (FIA + LC)
3	Fecarotta et al. (2025)	Italy (reference center)	Acylcarnitines/amino acids + 105-gene NGS → IMD	First-tier	Program evaluation (LC-MS/MS + NGS)
4	Grecsó et al. (2021)	Hungary (Szeged)	Steroids (17-OHP, 21-deoxycortisol, etc.) → CAH	Second-tier (QC)	QC-material method comparison
5	Lawrence et al. (2026)	Canada (NSO)	21-deoxycortisol + steroid profile → SW-CAH	Second-tier	Diagnostic-performance / algorithm
6	Hong et al. (2018)	USA (Univ. Washington)	C26:0-LPC, GALT, biotinidase → X-ALD, galactosemia, biotinidase deficiency	First-tier	Method development (multiplex)
7	Hubbard et al. (2009)	USA (multi-site)	C26:0-lysoPC → X-ALD	First-tier	Method development / validation
8	Kilgore et al. (2023)	USA (CDC)	Amino acids, LPC, organic acids → 10 disorders	Second-tier	Method development / validation (HILIC)
9	Klippel et al. (2025)	USA (Washington State)	Proteotypic peptides → Wilson disease, inborn errors of immunity	First-tier (novel)	Method development (proteomics)
10	la Marca et al. (2008)	Italy (Tuscany)	Acylcarnitines/amino acids → >40 IEM	First-/second-tier	Program evaluation (6-year)
11	Mak et al. (2023)	USA (California NBS)	121 metabolites → GA1, MMA, OTCD, VLCADD	Second-tier	Validation + comparison + ML (Random Forest)
12	Menkovic et al. (2019)	Canada (Sherbrooke)	HS, DS, creatinine → MPS	Population (urine)	Method development / reference values (UPLC-MS/MS, dried urine)
13	Engström et al. (2026)	Sweden (Karolinska, national)	Steroid ratio + 21-deoxycortisol → CAH	Second-tier	Program evaluation

Cont. Table 3

#	Study (Year)	Country / program	Target analytes → disorders	Tier	Design / method type
14	Oboshi et al. (2026)	Japan (multicenter)	DS/HS/KS disaccharides → MPS I/II	Second-tier	Method development / accuracy
15	Oguni et al. (2020)	Japan (+ Vietnam, USA)	5-plex enzyme activities → MPS I/II/IVA/VI/VII	First-tier	Analytical validation (enzyme assay)
16	Pajares-García et al. (2024)	Spain (Catalonia)	46 metabolites (OA, acylglycines, acylcarnitines, Hcy, orotic acid) → IEM	Second-tier	Method development / validation (underivatized)
17	Phoungphosop et al. (2026)	Thailand (national CH)	37-peptide barcode (12 proteins) → congenital hypothyroidism	Adjunct first-tier	Biomarker discovery (MALDI-TOF + LC-MS/MS)
18	Posern et al. (2024)	Germany (Hamburg) + USA	N-acetyl-L-aspartate → Canavan disease	First-tier	Method development
19	Abdulridha et al. (2024)	Iran (Mashhad/Tehran)	34 acylcarnitine derivatives → IEM	First-tier	Reference-value study
20	Saeed et al. (2024)	Iran (Mashhad)	Amino acids → metabolic disorders	First-tier	Reference-value study
21	Sidorina et al. (2026)	Italy (Rome, Bambino Gesù)	13 lipid biomarkers → GM1/GM2, Fabry, Gaucher, Krabbe, ASMD, NPC, X-ALD, PBD, MLD, MEDNIK	First-/second-tier / confirmatory	Method / clinical validation
22	Stapleton et al. (2020)	USA (+ Brazil, Japan)	GAGs (HS, DS) → MPS	First-/second-tier	Method development
23	Wang et al. (2013)	China	24 underivatized amino acids → metabolic diseases	Second-tier	Method development
24	Xie et al. (2025)	USA (California NBS)	Targeted metabolomics → GA-I, PA/MMA, OTCD, VLCADD	Second-tier	Method comparison (+ genome sequencing + AI/ML)

ASD = autism spectrum disorder; ASMD = acid sphingomyelinase deficiency; CAH = congenital adrenal hyperplasia; CH = congenital hypothyroidism; DBS = dried blood spot; DS = dermatan sulfate; GAG = glycosaminoglycan; HILIC = hydrophilic-interaction liquid chromatography; HS = heparan sulfate; IEM/IMD = inborn errors of metabolism / inherited metabolic disease; KS = keratan sulfate; LPC = lysophosphatidylcholine; MPS = mucopolysaccharidosis; NBS = newborn screening; NGS = next-generation sequencing; X-ALD = X-linked adrenoleukodystrophy.

3.3 Analytical challenges

The analytical challenges described in the studies included were categorized into three interrelated groups summarized in Table 4: preanalytical and dried-blood-spot matrix effects, analytical and instrumental challenges during the measurement, and calibration, cutoff derivation, and quantification challenges. The categories are not mutually exclusive, as errors that occur at an earlier stage often cascade through the stages. The following subsections discuss each category in turn and reference to the specific studies in which it was documented.

3.3.1 Pre-analytical and DBS matrix effects

Pre-analytical variability was consistently attributed to the dried-blood-spot specimen itself. The hematocrit is known to be one of the most important factors for blood spreading and for the volume of blood in a fixed diameter punch, so that quantitative results could be affected if the hematocrit of the specimen differs from the hematocrit of the calibrators, and analyte stability is also affected by the ambient humidity and temperature during the drying and storage process (De Kesel et al., 2013). Some additional minor variation may be introduced by the position of the punch in the spot, and most of the studies included here used dried blood spots with a fixed punch, which were thus subject to these effects.

Other biological and contamination related confounder data were also documented. Fetal glycosaminoglycan

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

contamination was found to be a confounder in the measurement of glycosaminoglycans for the mucopolysaccharidoses and required careful definition of cutoff and second tier re-screening (Stapleton et al., 2020). The metabolite signal was found to be highly dependent on gestational age, age at sampling and month of birth, all of which made it difficult to differentiate between cases and controls in the untargeted metabolomic profiling of newborn dried blood spots (Courraud et al., 2021).

Another pre-analytical source of variability was the preparation of quality control materials. Five in-house preparation methods were tested for the second tier of screening for congenital adrenal hyperplasia (CAH) and the one that involved repeated freeze-thaw cycles resulted in steroid concentrations that were significantly lower than the nominal values and higher than the acceptance limits (Grecsó et al., 2021). These discrepancies, which depend on the preparation, directly impact on the comparability of laboratory developed assays.

3.3.2 Analytical and instrumental challenges

One of the many analytical problems that recurred was the resolution of isobar and isomer species which are not separated by flow-injection tandem mass spectrometry. For instance, leucine and isoleucine both have a transition m/z 132 to 86, which means that they cannot be separated by MS alone; in this case chromatographic separation was needed to separate the two species by retention time (Wang et al., 2013).

It was demonstrated that there are different trade-offs between the choice of derivatized and underivatized protocols. The underivatized methods are simpler, but the use of ion pairing reagents may cause ion suppression in the ion source and decrease the column working life, while the derivatization with butyl-ester introduces more preparation steps, which can lead to the formation of artifacts and loss of quantification stability; the use of volatile modifiers like acetic acid was reported to reduce the ion suppression (Wang et al., 2013). The post-extraction addition strategy

Table 4. Analytical challenges in LC-MS/MS-based neonatal screening, by category, with representative evidence and reported mitigation

Category	Specific challenge	Mechanism / manifestation	Representative studies	Reported mitigation
Pre-analytical / DBS matrix	Hematocrit effect	Hematocrit governs blood spreading and the volume contained within a fixed-diameter punch, biasing quantification when specimen hematocrit departs from that of the calibrators	De Kesel et al. (2013)	Volumetric absorptive microsampling; hematocrit-matched calibration; whole-spot analysis
	Analyte stability	Degradation of labile analytes under elevated humidity and temperature during drying and storage	De Kesel et al. (2013)	Controlled drying; low-humidity, low-temperature storage
	Spot inhomogeneity / punch location	Radial chromatographic spreading within the spot; central-versus-peripheral punch variation	De Kesel et al. (2013)	Standardized central punch; defined spotting volume
	Biological contamination — fetal glycosaminoglycans	Fetal glycosaminoglycans elevate background and confound disease-specific glycosaminoglycan markers for the mucopolysaccharidoses	Stapleton et al. (2020)	Percentile-based cutoffs; second-tier re-screening
	Physiological confounders	Gestational age, age at sampling, and month of birth distort untargeted metabolite signals, obscuring case–control separation	Courraud et al. (2021)	Covariate adjustment; matched controls
	Exogenous contamination — pivalate	Pivalic acid arising from antibiotic administration or skin-emollient application elevates C5-acylcarnitine isomers	Pajares-García et al. (2024)	Isomer-resolving chromatography; isomer-specific second-tier markers
	Quality-control material preparation	In-house preparation dependent; repeated freeze–thaw spiking yielded steroid concentrations significantly below nominal values (exceeding ±15%)	Grecsó et al. (2021)	Standardized QC-preparation protocol; commutable materials

Cont. Table 4

Category	Specific challenge	Mechanism / manifestation	Representative studies	Reported mitigation
Analytical / instrumental	Matrix effects and ion suppression/ enhancement	Co-eluting phospholipids, salts, and buffer additives suppress or enhance ionization in the source	Matuszewski et al. (2003); Wang et al. (2013); Courtney et al. (2026)	Post-extraction matrix-effect assessment; volatile modifiers; chromatographic clean-up
	Isobaric/isomeric amino acids	Leucine, isoleucine, and alloisoleucine share the transition m/z 132 → 86; hydroxyproline and α-alanine/sarcosine interfere	Wang et al. (2013); Mak et al. (2023)	Chromatographic separation by retention time
	Isobaric acylcarnitine isomers	C4 (butyryl/isobutyryl) and C5 (isovaleryl/ valeryl; pivaloyl/2-methylbutyryl) isomers are indistinguishable by MRM alone	Mak et al. (2023); Pajares-García et al. (2024)	LC or ion-mobility high-resolution MS; isomer-specific markers
	Derivatized vs underivatized workflow	Butyl-ester derivatization adds steps and artifacts; underivatized methods may require ion-pairing reagents that cause ion suppression and shorten column life	Wang et al. (2013); Pajares-García et al. (2024)	Volatile modifiers (acetic acid); validated underivatized multiplex panels
	Internal-standard availability and cost	Stable-isotope-labeled internal standards improve accuracy but are costly and not universally available; some methods used non-deuterated standards	Wang et al. (2013)	Stable-isotope-labeled internal standards where obtainable
	Carryover / source contamination	Sample-to-sample carryover and source fouling degrade quantitative reliability	Assessed per Clinical and Institute (2014)	Wash cycles; carryover verification within method validation
Calibration / cutoffs / quantification	Cutoff derivation	Population-percentile cutoffs (for example, the 95th percentile) are sensitive to the reference population	Stapleton et al. (2020)	Locally validated percentile cutoffs
	Reference-range localization	Locally derived acylcarnitine and amino-acid ranges diverge from widely used global (R4S) values	Abdulridha et al. (2024); Saeed et al. (2024)	Population-specific reference intervals
	Covariate-dependent cutoffs	Steroid cutoffs for congenital adrenal hyperplasia require adjustment for gestational age, birth weight, and age at sampling	Lawrence et al. (2026); Engström et al. (2026)	Covariate-adjusted interpretation
	Analyte ratios versus absolute concentration	Absolute concentrations are insufficient for specificity; ratios and dual-marker rules improve discrimination	Lawrence et al. (2026); Oboshi et al. (2026)	Ratio-based and dual-marker decision rules
	Semi-quantitative FIA versus quantitative LC	Flow-injection analysis lacks chromatographic separation; LC provides isomer-resolved quantitative output	Wang et al. (2013); Kilgore et al. (2023)	LC-MS/MS second-tier quantification
	Absence of commutable calibrators	The scarcity of commutable certified calibrators drives inter-laboratory discordance	Synthesis (see Section 3.4.4)	Commutable certified reference materials; IDMS traceability

DBS = dried blood spot; FIA = flow-injection analysis; IDMS = isotope-dilution mass spectrometry; MRM = multiple-reaction monitoring; R4S = Region 4 Stork. Total parenteral nutrition and EDTA anticoagulant are additional recognized exogenous contamination sources, although they were not specifically documented within the included studies.

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

(Matuszewski et al., 2003) was used to characterize matrix effects and broad underivatized multiplex second tier panels were validated (Pajares-García et al., 2024).

Other instrumental factors included multiplexing, throughput and reagent complexity. A universal second-tier assay based on hydrophilic-interaction liquid chromatography was developed for the simultaneous injection of amino acids, lysophosphatidylcholines, and organic acids for ten disorders (Kilgore et al., 2023), while a five-plex enzyme-activity assay for the mucopolysaccharidoses was limited by reagent complexity (Oguni et al., 2020). The reported analytical time was from a long analytical run of 17 to 19 minutes (Wang et al., 2013) to a one-minute ultra-high performance method (Menkovic et al., 2019) which is the constant compromise between analytical breadth and throughput.

3.3.3 Calibration, cutoffs, and quantification

Derivation and quantification of the cutoff emerged as key factors in the screening performance. Percentiles of a newborn control population were often used as cutoffs as in the case of glycosaminoglycan markers, which were defined as the 95th percentile (Stapleton et al., 2020), but these were found to be sensitive to local factors. Acylcarnitines (Abdulridha et al., 2024) and amino acids (Saeed et al., 2024) reference ranges were demonstrated to be different from the commonly used reference ranges across the globe, and the reference range for amino acids and the steroid cut-off for congenital adrenal hyperplasia (CAH) were found to need adjustment for gestational age, birth weight, and age at sampling to control the false-positive rate.

Analyte ratios were repeatedly reported as useful for increasing specificity when absolute concentrations were inadequate. The use of steroid-ratio algorithms with 21-deoxycortisol significantly enhanced the positive predictive value of the screening test for congenital adrenal hyperplasia (Engström et al., 2026; Lawrence et al., 2026); and dual-marker decision rules led to an increase in the specificity for the mucopolysaccharidoses (Oboshi et al., 2026). In the entire corpus, the lack of availability of interchangeable certified calibrators was a major limitation for the comparability of quantitative results from different laboratories.

3.4 Standardization strategies

The standardization approaches that were identified in the included studies were found to be implemented at four complementary levels (Table 5): external quality assessment and proficiency testing, as in the case of developing a dried-blood-spot (DBS) proficiency testing materials for the lysosomal disorders (Courtney et al., 2026); consensus guidelines and laboratory accreditation; post-analytical data harmonization via covariate-adjusted interpretive tools; and metrological traceability, which involves reference measurement procedures and certified reference materials. Each level is examined in the subsections that follow.

3.4.1 External quality assessment and proficiency testing

External quality assessment and proficiency testing were cited as the most important means of monitoring inter-laboratory consistency, and the Centers for Disease Control and Prevention (CDC) Newborn Screening Quality Assurance Program was mentioned most often. One of the most significant advances in the corpus was the creation of first-tier dried-blood-spot (DBS) proficiency-testing materials for lysosomal diseases using recombinant enzymes, expanding the panel to seven conditions, with mucopolysaccharidosis type II added, and showed acceptable performance on flow-injection tandem mass spectrometry, liquid-chromatography tandem mass spectrometry, and the digital microfluidics and fluorometric platforms (Courtney et al., 2026).

The harmonizing effect of such schemes has been proven, as use of common reference materials via the program was reported to decrease the variation of results obtained for proficiency testing among participating laboratories (Pickens et al., 2020), following many decades of quality-assurance activities (De Jesús et al., 2015). The recombinant-enzyme approach further improved scalability and enabled international enrollment, although genuine convergence of routine results across all participating programs was reported to remain incomplete (Courtney et al., 2026).

3.4.2 Guidelines, consensus documents, and accreditation

A second level of standardization was consensus guidelines and formal accreditation. The consensus documents which outline best practice in tandem mass spectrometry for newborn screening and in development and verification of liquid chromatography–mass spectrometry methods (Clinical & Institute, 2014; Hoffman, 2010) and accreditation to recognised laboratory-quality standards such as ISO 15189 support procedural harmonisation. The instruments establish expectations for calibration, quality control, interference evaluation, acceptance of the run, and

Table 5. Standardization strategies for LC-MS/MS-based neonatal screening, by level, with reported effect or status

Level	Mechanism / example	Representative studies	Reported effect / status
External quality assessment and proficiency testing	CDC Newborn Screening Quality Assurance Program (NSQAP): distribution of quality-control and proficiency-testing DBS materials; recombinant-enzyme first-tier lysosomal-disease panel expanded to seven conditions and evaluated across FIA-MS/MS, LC-MS/MS, digital microfluidics, and fluorometry	Courtney et al. (2026); Pickens et al. (2020); De Jesús et al. (2015)	Common reference materials reduced the dispersion of proficiency-testing results between laboratories; scalable materials enabled international enrollment; full convergence of routine results remains incomplete
	ERNDIM and national EQA schemes: inter-laboratory comparison for inherited metabolic disorders	Recognized schemes; not represented among the included studies	Provide external comparison; effect not directly quantified within the corpus
Guidelines, consensus documents, and accreditation	CLSI consensus documents for tandem mass spectrometry in newborn screening and for LC-MS methods: define calibration, quality control, interference assessment, and verification	Hoffman (2010); Clinical and Institute (2014)	Procedural harmonization; explicit adherence inconsistently reported
	CLSI guidance for defining and verifying reference intervals	Abdulridha et al. (2024); Saeed et al. (2024)	Locally established and verified reference intervals
	ISO 15189 laboratory-quality accreditation	General standard	Framework for demonstrated competence; accreditation status seldom reported in the corpus
Post-analytical data harmonization	Region 4 Stork (R4S) and its successor CLIR: multivariate, covariate-adjusted percentile tools adjusting for age, gestational age, and birth weight	McHugh et al. (2011)	Reduced false-positive rates without loss of sensitivity
	Analyte-ratio algorithms: steroid ratios with 21-deoxycortisol; dual-marker rules for the mucopolysaccharidoses	Lawrence et al. (2026); Engström et al. (2026); Oboshi et al. (2026)	Positive predictive value increased from 14% to 84% and false positives reduced by 88% for congenital adrenal hyperplasia; improved specificity for the mucopolysaccharidoses
	Machine-learning classifiers: random-forest and integrated AI/ML interpretation	Mak et al. (2023); Xie et al. (2025)	False-positive reduction at full sensitivity; false positives reduced by 98.8% with metabolomics, genome sequencing, and machine learning combined
Metrological traceability and reference materials	Recombinant-enzyme DBS reference and quality-control materials: scalable alternative to patient-cell-line preparations, characterized across multiple platforms	Courtney et al. (2026)	Improved scalability and reproducibility of reference materials
	Stable-isotope-labeled internal standards for quantitative traceability	Corpus-wide; non-deuterated internal standards noted as a limitation in Wang et al. (2013)	Partial traceability; constrained where non-deuterated standards are used
	Commutable certified reference materials and isotope-dilution reference measurement procedures	Gap; largely absent within the corpus	Least-developed layer; alignment of routine measurements with reference procedures remains incomplete

CLIR = Collaborative Laboratory Integrated Reports; DBS = dried blood spot; EQA = external quality assessment; ERNDIM = European Research Network for evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism; FIA-MS/MS = flow-injection-analysis tandem mass spectrometry; NSQAP = Newborn Screening Quality Assurance Program; R4S = Region 4 Stork.

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

verification of the method which are directly applicable to screening using LC-MS/MS.

It was found that the extent of explicit reporting of adherence was different across the corpus. In the reference-value studies, population-specific reference intervals were developed and confirmed following the recognized guidelines for the development and confirmation of reference intervals (Abdulridha et al., 2024; Saeed et al., 2024) and a number of method-development reports were validated against analytical-performance criteria. However, there was not a uniform reporting of explicit documentation of adherence to the guidelines, and of accreditation status, limiting the assessment of procedural standardization across programs and limiting the comparability of laboratory-developed assays.

3.4.3 Post-analytical data harmonization

Post-analytical data harmonization was represented primarily by multivariate, covariate-adjusted interpretive tools, the most important of which being the worldwide collaborative project that developed clinically valid target ranges for the interpretation of results, and the database supporting this (McHugh et al., 2011). This reliance on population-specific characteristics was confirmed within the corpus: population-specific reference ranges for acylcarnitines and amino acids were found to be different from the global reference ranges commonly used, and the cutoffs for steroid diagnosis of congenital adrenal hyperplasia (CAH) were shown to vary according to gestational age, birth weight, and age at sampling.

A new development in post-analytical harmonization was the use of machine-learning classifiers to lower the number of false-positive referrals, while maintaining high sensitivity. A targeted metabolomic panel, combined with a random-forest model, reduced the number of false positive results at full sensitivity for several disorders (Mak et al., 2023), and the combination of targeted metabolomics, genome sequencing and machine learning resulted in 98.8% fewer false positive results at full sensitivity (Xie et al., 2025). These methods are extensions of the covariate adjusted approach of previous interpretive tools to higher dimensional data.

3.4.4 Metrological traceability and reference materials

Metrological traceability and the availability of reference materials were addressed less frequently, yet several relevant advances were identified. To prepare materials for recombinant-enzyme dried-blood-spots, an alternative that was scalable and reproducible to those based on patient cell lines was developed, and these materials were characterized in several analytical platforms (Courtney et al., 2026). The most commonly used methods were all based on the use of a stable-isotope labelled internal standard to ensure accuracy and traceability for quantitative measurement.

However, there were still few certified reference materials available that could be used for the calibration of the methods, and some methods used non-deuterated internal standards (Wang et al., 2013), limiting the possibility of full metrological traceability. Thus, there was an incomplete reporting of alignment of routine screening measurements to reference measurement procedures throughout the corpus.

3.5 Future directions reported in primary studies

The use of higher resolution and more flexible instrumentation was a recurring suggestion. Untargeted and high-resolution metabolomic profiling of newborn dried blood spots was performed to interrogate large numbers of features and to enable retrospective analysis (Courraud et al., 2021; Xie et al., 2025), and multiplexed proteomic and peptide-based assays using liquid chromatography–tandem mass spectrometry and MALDI-TOF profiling were developed (Klippel et al., 2025; Phoungphosop et al., 2026). Digital microfluidics was also used for the preparation and evaluation of screening materials (Courtney et al., 2026), and volumetric absorptive microsampling was mentioned as an alternative to the traditional DBS that does not require hematocrit (De Kesel et al., 2013).

The expansion of the screening panel constituted a second theme. In addition to existing metabolic disorders, methods were reported for the mucopolysaccharidoses and other lysosomal disorders (Menkovic et al., 2019; Oboshi et al., 2026; Oguni et al., 2020; Stapleton et al., 2020) and for X-linked adrenoleukodystrophy, based on the measurement of C26:0-lysophosphatidylcholine (Hong et al., 2018; Hubbard et al., 2009), and for congenital adrenal hyperplasia, second-tier steroid profiling (Engström et al., 2026; Lawrence et al., 2026). Single condition targets were added for newer conditions such as Wilson disease, inborn errors of immunity (Klippel et al., 2025), Canavan disease (Posern et al., 2024), and inherited neurodegenerative disorders identified by lipid biomarkers (Sidorina et al., 2026).

The third direction was the biochemical screening combined with genomic technologies. The parallel use of

LC-MS/MS and next-generation sequencing successfully solved all the cases and led to a sequencing result in 37.9% of the cases (Fecarotta et al., 2025); targeted metabolomics in conjunction with genome sequencing was suggested to enhance the sensitivity and specificity of the test (Xie et al., 2025). Mass spectrometry was thus put in a functional position to complement the genomic results in a multi-omic workflow.

The fourth theme was the use of data-science techniques for post analytical interpretation. The logic of the existing covariate-adjusted interpretive tools (McHugh et al., 2011) was extended by training supervised machine-learning models to distinguish true positives from false positives and reduce the number of referrals while keeping full sensitivity (Mak et al., 2023; Xie et al., 2025). The need for rigorous external validation, transparent reporting, and appropriate data governance was identified as a prerequisite for routine deployment.

3.6 Risk of bias across the evidence base

The 24 studies included were subjected to design-stratified appraisal. QUADAS-2 (Table 6) was used to appraise diagnostic-accuracy and method-comparison studies, a fit-for-purpose validation framework adapted from the AOAC classification matrix (Table 7) was used to appraise method-development and reference-value studies, and diagnostic-accuracy reports, proficiency-testing (PT) reports, and external-quality-assessment (EQA) reports were appraised narratively against international PT and newborn-screening quality standards (Table 8). Ratings reflect reviewer judgment from the available texts; final signalling-question scoring requires full-text extraction in duplicate.

Per domain risk of bias (RoB) rated Low / High / Unclear; applicability rated for the first three domains. Signalling questions are given in the QUADAS-2 format (Patient selection: consecutive/random sampling, case–control avoided, no inappropriate exclusions; Index test: blinded interpretation, pre-specified threshold; Reference standard: correct classification, blinded interpretation, and flow and timing: appropriate interval, complete and uniform verification, all patients analysed).

Table 6. QUADAS-2 risk-of-bias and applicability assessment (diagnostic-accuracy / method-comparison studies)

Study (Year)	Patient selection	Index test	Reference standard	Flow & timing	Applicability	Dominant concern
Mak et al. (2023)	High	Unclear	Low	High	Low	Retrospective, case-enriched screen-positive set; class imbalance; incomplete verification of screen-negatives; ML threshold optimism
Oboshi et al. (2026)	High	Unclear	Low	High	Low–moderate	Retrospective 119-case set; combined metabolomics/ML/genome-sequencing thresholds; partial verification
Oboshi et al. (2026)	High	High	Low	Unclear	Low	Case–control structure; dual-marker cut-off derived post hoc; few confirmed cases
Lawrence et al. (2026)	High	Unclear	Low	Unclear	Moderate	Monte-Carlo simulation rather than consecutive prospective enrolment; modelled thresholds
Sidorina et al. (2026)	High	Unclear	Low	Unclear	Moderate	Case-confirmation design (89 patients + controls); mixed plasma/DBS; diagnostic rather than population screening
Fecarotta et al. (2025)	Unclear	Low–unclear	Low	Unclear	Low	Prospective program cohort but modest positive-case number; screen-negatives not fully verified

The areas with the lowest accuracy scores in the accuracy studies are patient selection and flow-and-timing, due to case enrichment, retrospective sampling and differential/partial verification of screen-negative neonates. Reference standards (molecular, enzymatic) are generally good; index-test blinding and threshold pre-specification is often ambiguous.

Framework adapted from the AOAC fit-for-purpose

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

classification matrix: the minimum validation level is matched to the intended purpose. Newborn screening is a regulatory/commercial screening application, for which the matrix specifies Harmonized Collaborative Validation (HCV) — full collaborative, multi-laboratory study preceded by successful single-lab (SLV) or multi-lab (MLV) validation. Reference-value studies are considered indicator/monitoring tests, in which SLV is the

Table 7. Fit-for-purpose validation appraisal (method-development and reference-value studies)

Study (Year)	Intended purpose (AOAC category)	Fit-for-purpose minimum	Level achieved	Validation elements emphasised	Fit-for-purpose gap
Kilgore et al. (2023)	Regulatory/commercial screening (universal 2TT)	HCV	SLV	LOD/LOQ, linearity, precision, selectivity across 10 disorders	Single-lab only; no MLV/HCV
Hong et al. (2018)	Regulatory/commercial screening (multiplex 1st-tier)	HCV	SLV	Linearity, precision, selectivity (multiplex)	Single-lab only
Hubbard et al. (2009)	Regulatory/commercial screening (X-ALD 1st-tier)	HCV	SLV	Precision, recovery, selectivity; multi-site sample sourcing	Validation not collaborative
Klippel et al. (2025)	Screening (novel proteomic assay)	HCV (pre-implementation)	SLV (early-stage)	Proof-of-concept selectivity, precision	Discovery/early validation; MLV–HCV pending
Posern et al. (2024)	Screening (Canavan 1st-tier)	HCV (pre-implementation)	SLV	LOD/LOQ, linearity, precision	Pre-screening single-lab method
Oguni et al. (2020)	Regulatory/commercial screening (5-plex enzyme)	HCV	SLV (international collaborators)	Precision, linearity, selectivity	Collaborators involved but not formal MLV/HCV
Stapleton et al. (2020)	Screening (GAG 1st-/2nd-tier)	HCV	SLV	Precision, recovery; fetal-GAG confounder characterised	Single-lab; contamination confounder
Wang et al. (2013)	Regulatory/commercial screening (2TT amino acids)	HCV	SLV	Linearity, precision, selectivity (underivatised)	Single-lab; long run time
Pajares-García et al. (2024)	Regulatory/commercial screening (2TT panel, operational)	HCV	SLV (single program)	LOD/LOQ, precision, linearity, program data	Single-program validation
Menkovic et al. (2019)	Screening (urine MPS) / reference values	HCV	SLV	Precision (RSD <9%), bias (<7%), reference values (n = 500)	Single-lab; no diseased-cohort accuracy
Abdulridha et al. (2024)	Indicator/monitoring (reference ranges)	SLV	SLV	Reference ranges for 34 acylcarnitines vs external labs	Population-specific; no inter-lab comparison
Saeed et al. (2024)	Indicator/monitoring (reference ranges)	SLV	SLV	Amino-acid reference ranges (n = 1,005) vs R4S	Single-region; population-specific
Grecsó et al. (2021)	Process/quality (QC-material preparation)	SLV–MLV (standardisation)	SLV	Comparison of DBS QC preparation techniques	Single-lab QC study
Courraud et al. (2021)	Research/discovery (untargeted)	Not applicable (exploratory)	SLV-level exploratory	Feature detection; confounder analysis	Discovery only; no fit-for-purpose validation

reference value. Validation levels: SLV (single-laboratory: inclusivity/exclusivity [selectivity], ruggedness, stability, lot-to-lot, and for quantitative methods LOD, LOQ, RSDr, linearity vs a reference method), MLV (≥2 laboratories), HCV (full collaborative study, robust statistics without outlier removal except for assignable cause).

The dominant, systematic gap is that most methods intended for a regulatory screening purpose have only reached single-laboratory validation, whereas the fit-for-purpose standard for that purpose is collaborative multi-laboratory (HCV) validation. This is the analytical counterpart to the single-center design limitation and is the strongest recurrent risk-of-bias signal in the corpus.

The domains of appraisal relate to the governing standards. Cells: A = adequately addressed, P = partially addressed, NR = not reported, NA = not applicable to the study type.

Table 8. Narrative methodological appraisal — EQA/PT and program-evaluation items

Appraisal domain (governing standard)	**Courtney (2026) PT materials**	**Grecsó (2021) QC materials**	**la Marca (2008) program**	**Engström/ Nordenström (2026) program**	**Fecarotta (2025) program**
PT-provider competence & scheme design (ISO/IEC 17043:2023)	A (CDC NSQAP scheme)	P (single-lab QC, not a PT provider)	NA	NA	NA
Statistical design, assigned values & robust/outlier handling (ISO 13528:2022; ISO 5725)	P	P	NA	NA	NA
Metrological traceability & commutability of materials (ISO 17511; CDC NSQAP)	A (recombinant-enzyme characterisation)	P	NA	NA	NA
Analytical acceptance criteria per CLSI newborn-screening guidance	P	P	P (cut-offs, false-test reduction)	A (2TT cut-offs)	P
Program quality indicators — PPV, false-positive rate, detection/incidence (Yusuf et al., 2019)	NA	NA	P (FP/FN reduction described)	A (PPV 14%→84%; FP −88%)	P (case resolution; FP characterised)
Timeliness indicators (Yusuf et al., 2019), QI 5a–5f)	NA	NA	NR	NR	NR
Quality-improvement / evaluation loop & transparency of limitations (CDC program-evaluation framework)	P	P	A	A	P

Program-evaluation reports reasonably well document diagnostic yield, but do not generally report on the reduction of false-positive results, and do not usually report on the statistical scheme design, the assignment of the values, or the commutability expected under ISO/IEC 17043 and ISO 13528 for EQA/PT and QC-material studies.

4. Discussion

4.1 Summary of principal findings

The synthesis focused on three questions regarding the analytical challenges, the standardization strategies and the future directions of LC-MS/MS in neonatal screening. With regards to the first question, the main challenges were pre-analytical, which were related to the dried-blood-spot matrix and to biological and contamination related confounders; analytical, related to isobaric and isomeric interferences, requiring chromatographic separation; and quantitative, because of the population dependent cutoffs, and the lack of commutable certified calibrators (De Kesel et al., 2013; Stapleton et al., 2020; Wang et al., 2013).

As for the second and third questions, the most strongly

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

endorsed strategies were the use of post-analytical, covariate-adjusted interpretive tools and the use of analyte-ratio algorithms, in addition to expanding the materials used for proficiency testing (Courtney et al., 2026; Lawrence et al., 2026; McHugh et al., 2011). The future directions reported were all focused on biochemical screening and genomic sequencing and machine learning (Fecarotta et al., 2025; Mak et al., 2023; Xie et al., 2025). Analytical capability was therefore identified as being mature, while harmonization of the laboratories and programs was identified as the main constraint.

4.2 Interpreting the analytical challenges

The analytical problems that were found in this review should be considered as interdependent rather than separate problems, because a problem at one stage is often carried through to the next. The lower bound of achievable accuracy is set by pre-analytical variability, since the spreading of blood depends on the hematocrit and the amount of blood collected from a standard-sized punch depends on the diameter of the punch, and the stability of the analyte depends on the drying and storage conditions (De Kesel et al., 2013). Fetal glycosaminoglycans which interfere with the measurement of disease-specific glycosaminoglycans, also have contamination-related effects (Stapleton et al., 2020).

One of the most important features of liquid chromatography is its ability to separate isobaric and isomeric species, which cannot be separated by flow-injection tandem mass spectrometry, such as leucine and isoleucine, which have the same multiple-reaction-monitoring transition and are otherwise indistinguishable (Wang et al., 2013). This is very important for specificity but is not a solution to the upstream/downstream limitations of separation by the matrix and by the metrology. Chromatography does not solve the problem of variability due to in-house quality-control preparation (Grecsó et al., 2021) nor the lack of commutable calibrators, which results in population-specific reference ranges that differ from the global ones (Abdulridha et al., 2024; Saeed et al., 2024).

The effects of these lingering problems on the clinical outcomes are significant. Enzyme-activity screening without a specific second-tier test results in a high false-positive rate (FP) in part due to the misclassification of carriers and pseudodeficiency alleles, which can lead to unnecessary confirmatory testing of unaffected infants, and create logistical and psychological burden for families and health systems (Oboshi et al., 2026). This burden can be further reduced by applying second tier LC-MS/MS, of analyte-ratio decision rules, and of machine-learning classifiers, without compromising sensitivity (la Marca et al., 2008; Mak et al., 2023; Xie et al., 2025).

4.3 The path to harmonization

The evidence shows that harmonization cannot be achieved with a single intervention and it is an outcome of a coordinated action of a few complementary layers. Each of these layers tackles a different source of inter-laboratory discordance: external quality assessment, consensus guidelines and accreditation, post-analytical interpretive tools and metrological traceability.

Post-analytical levers were identified as the most readily available with minimal need for re-tooling of the analytical process as they leverage on data that laboratories already collect. The use of covariate-adjusted interpretive tools that adjust for age, gestational age, and birth weight significantly improves the specificity of screening without compromising sensitivity (McHugh et al., 2011) and analyte-ratio algorithms provide similar benefits, as evidenced by the dramatic improvement in the positive predictive value of the screening for congenital adrenal hyperplasia when 21-deoxycortisol and steroid ratios are added (Engström et al., 2026; Lawrence et al., 2026).

Upstream harmonization depends upon shared reference materials and adherence to consensus standards. Consensus guidance and accreditation establish common procedural expectations (Clinical & Institute, 2014; Hoffman, 2010), and the dispersion of results between laboratories is minimized by the distribution of common proficiency-testing materials (Pickens et al., 2020), and production of materials containing recombinant enzymes has been scaled up to a global level (Courtney et al., 2026).

The least developed layer is metrological, because there is still a lack of commutable certified reference materials and even traceability to reference measurement procedures. The benefits of the post-analytical adjustment are therefore likely to need a continued investment in this final layer, so that the measurement itself is tied to a common metrological reference, which supports the post-analytical adjustment.

4.4 Future directions and translational implications

The directions of future research identified in the primary studies have different translational implications when compared to the limitations of routine screening. While high-resolution and untargeted metabolomic profiling allows interrogation of large feature sets and

allows the re-analysis of archived specimens (Courraud et al., 2021; Xie et al., 2025), this ability also increases the amount of data to be interpreted and validated, and the reliance on pre-analytical covariates must be controlled before the technology can be used in a prospective manner. The use of proteomics and peptides for multiplexing allows screening of diseases without informative small molecule markers, including congenital hypothyroidism and Wilson disease (Klippel et al., 2025; Phoungphosop et al., 2026). Coupling biochemical screening with genomic sequencing puts mass spectrometry in the position of being a functional confirmatory measurement (Fecarotta et al., 2025), but introduces the new challenges of cost, variants of uncertain significance and consent issues that need to be addressed before widespread adoption.

While significant decreases in false-positive referrals can be achieved with machine learning applied to post-analytical interpretation, this requires robust external validation, reporting on the models, and proper data governance. Other sampling methods like volumetric absorptive microsampling can reduce the hematocrit effect (De Kesel et al., 2013) but these methods need to be method transferred and revalidated before they can be used in a screening program.

In all these directions, the ultimate restriction on the uptake of translation is throughput, cost, regulatory validation and accreditation. The overwhelming presence of single-center studies in the corpus highlights the need for multi-site reproducibility before any of the innovations can be scaled up and for implementation science to be as much of a factor as analytical performance in the move from proof of concept to routine practice.

4.5 Global health equity implications

The location of the evidence is strongly suggestive of global health equity. The studies that were found were mostly from high income countries, and while two studies from a middle-income country were found, the majority of low and middle-income countries were not represented in the corpus. This focus means that the results are not generalizable and are a reflection of greater inequity in access to LC-MS/MS-based screening.

The reference-value studies highlighted the need for locally derived reference values for acylcarnitines and amino acids, which differed from the global reference values (Abdulridha et al., 2024; Saeed et al., 2024). The adoption of the global cutoffs without critical evaluation by the under-represented population may therefore result in greater misclassification and make the setting of local reference intervals an equity issue rather than a technical one.

Several routes toward greater equity were nevertheless apparent. The design of methods that will reduce the cost will make them more easily accessible to implement (Wang et al., 2013), and the scalable production of proficiency-testing materials has already led to international participation in quality-assurance programs (Courtney et al., 2026). Sustained investment in laboratory infrastructure, cold-chain logistics, reagent supply, and workforce training nonetheless remains a prerequisite, since analytical methods alone cannot close a gap that is ultimately structural in nature.

4.6 Strengths and limitations

This review has a number of methodological advantages. The review was conducted and reported in accordance with the PRISMA 2020 statement (Page et al., 2021). The study selection and data extraction were done independently and in duplicate, the design-stratified approach to critical appraisal was used (Clinical & Institute, 2014; Hong et al., 2018; Whiting et al., 2011), and the synthesis was reported following the guideline of the Synthesis Without Meta-analysis (Campbell et al., 2020).

The need to acknowledge limitations due to the evidence base should also be recognized. The methodological and clinical diversity of the included studies did not allow quantitative meta-analysis; most of the studies were validations of single centers and/or single programs, with limited external reproducibility; and the matrix of specimens was not uniform, as dried urine and mixed plasma and dried-blood-spot designs were also represented (Menkovic et al., 2019; Sidorina et al., 2026).

There are also limitations of the review process acknowledged. The exclusion of publications in other languages may cause language bias, there may be publications which are not included, and the field is rapidly developing, so the most up-to-date methods may not be adequately represented. Furthermore, one record was kept as a borderline inclusion as the analytical target was a research instead of a screening application (Courraud et al., 2021) and the fact that it was not included in the synthesis was handled transparently so that the impact on the conclusions could be tracked.

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

5. Conclusion

The liquid chromatography–tandem mass spectrometry (LC–MS/MS) technique has been the cornerstone of current neonatal screening and the evidence presented in this review suggests that the analytical power of this technique is now developed. The platform can multiplex the quantification of amino acids and acylcarnitines, a growing collection of second tier tests, enzyme-activity and glycosaminoglycan analysis for the lysosomal disorders, steroid profiling for congenital adrenal hyperplasia, and novel proteomic and peptide-based assays.

Despite this maturity, there are still some challenges that remain. Pre-analytical variability from the dried-blood-spot matrix, interferences from isobars and isomers and population-dependent cutoffs, as well as the limited availability of commutable certified reference materials are still sources of inter-laboratory variability. The challenges are interrelated, so that specificity achieved at the analytical stage cannot offset error that may occur prior to or after measurement.

The overall recommendation that comes out is that coordination of the standardization of the platform in the pre-analytical, analytical and post-analytical domains is necessary to realize the full value of the platform. Proficiency-testing materials should be shared and consensus guidelines and accreditation should be followed, covariate-adjusted post-analytical interpretive tools should be implemented, and metrological traceability should be pursued in conjunction with the other elements (Courtney et al., 2026; McHugh et al., 2011). Of these, post analytical harmonization and analyte-ratio interpretation were identified as having the greatest potential for immediate benefits, while the development of commutable reference materials was the most significant unmet need.

The priority for future research is therefore the generation of multi-site reproducibility studies, the development of commutable certified reference materials, the prospective validation of genomic and machine-learning integration, and the deliberate inclusion of populations from low- and middle-income settings. Further development in this direction would lead to the development of harmonized and equitable harmonized screening performance of LC-MS/MS, and to the parallel development of technical innovation with harmonization and implementation.

Statements and Declarations

Conflict of Interest

The author declares no conflicts of interest or competing interests.

Funding

No funding was received.

Data Availability

This study is based on a systematic literature synthesis and does not involve the collection or analysis of primary or proprietary datasets. The data underlying this study consist of the 24 peer-reviewed journal articles identified through the PRISMA 2020-guided systematic search of four electronic databases, namely PubMed/MEDLINE, Scopus, Europe PMC, and the Cochrane Library.

Acknowledgement

None

References

Abdulridha, J. S., Mashkani, B., Alaei, A., Boskabadi, M., Varasteh, A., & Keyfi, F. (2024). Determination of normal range of acylcarnitine in neonatal dried blood spots using LC-MS/MS. Reports of Biochemistry & Molecular Biology, 12(4), 522. https://doi.org/10.61186/rbmb.12.4.522

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., & Thomas, J. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. bmj, 368. https://doi.org/10.1136/bmj.l6890

Chace, D. H., Millington, D., Terada, N., Kahler, S., Roe, C. R., & Hofman, L. (1993). Rapid diagnosis of phenylketonuria by quantitative analysis for phenylalanine and tyrosine in neonatal blood spots by tandem mass spectrometry. Clinical chemistry, 39(1), 66-71. https://doi.org/10.1093/clinchem/39.1.66

Clinical, & Institute, L. S. (2014). Liquid chromatography-mass spectrometry methods; approved guideline. CLSI document C62-A. In: CLSI Wayne (PA).

Courraud, J., Ernst, M., Svane Laursen, S., Hougaard, D. M., & Cohen, A. S. (2021). Studying autism using untargeted metabolomics in newborn screening samples. Journal of Molecular Neuroscience, 71(7), 1378-1393. https://doi.org/10.1007/s12031-020-01787-2

Courtney, E., Isenberg, S. L., Lim, T., Pickens, C. A., Lee, R., Cuthbert, C., & Petritis, K. (2026). Development of Dried Blood Spot Proficiency Testing Materials for Newborn Screening of Lysosomal Diseases Using Recombinant Enzymes. International journal of neonatal screening, 12(2), 40. https://doi.org/10.3390/ijns12020040

De Jesús, V. R., Mei, J. V., Cordovado, S. K., & Cuthbert, C. D. (2015). The newborn screening quality assurance program at the centers for disease control and prevention: thirty-five year experience assuring newborn screening laboratory quality. International journal of neonatal screening, 1(1), 13-26. https://doi.org/10.3390/ijns1010013

De Kesel, P. M., Sadones, N., Capiau, S., Lambert, W. E., & Stove, C. P. (2013). Hemato-critical issues in quantitative analysis of dried blood spots: challenges and solutions. Bioanalysis, 5(16), 2023-2041. https://doi.org/10.4155/bio.13.156

Engström, K., Zetterström, R. H., Wedell, A., & Nordenström, A. (2026). Neonatal Screening for CAH in Sweden-Results of Implementing Second-Tier Testing. International journal of neonatal screening, 12(2), 29. https://doi.org/10.3390/ijns12020029

Fecarotta, S., Vaccaro, L., Verde, A., Alagia, M., Rossi, A., Colantuono, C., Cacciapuoti, M. T., Annunziata, P., Riccardo, S., & Grimaldi, A. (2025). Combined biochemical profiling and DNA sequencing in the expanded newborn screening for inherited metabolic diseases: the experience in an Italian reference center. Orphanet Journal of Rare Diseases, 20(1), 38. https://doi.org/10.1186/s13023-025-03546-1

Grecsó, N., Zádori, A., Baráth, Á., Galla, Z., Rácz, G., Bereczki, C., & Monostori, P. (2021). Comparison of different preparation techniques of dried blood spot quality controls in newborn screening for congenital adrenal hyperplasia. Plos one, 16(5), e0252091. https://doi.org/10.1371/journal.pone.0252091

Guthrie, R., & Susi, A. (1963). A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. Pediatrics, 32(3), 338-343. https://doi.org/10.1542/peds.32.3.338

Hoffman, G. L. (2010). Newborn screening by tandem mass spectrometry: approved guideline. Clinical and Laboratory Standards Institute.

Hong, X., Kumar, A. B., Scott, C. R., & Gelb, M. H. (2018). Multiplex tandem mass spectrometry assay for newborn screening of X-linked adrenoleukodystrophy, biotinidase deficiency, and galactosemia with flexibility to assay other enzyme assays and biomarkers. Molecular genetics and metabolism, 124(2), 101-108. https://doi.org/10.1016/j.ymgme.2018.03.012

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

Hubbard, W. C., Moser, A. B., Liu, A. C., Jones, R. O., Steinberg, S. J., Lorey, F., Panny, S. R., Vogt Jr, R. F., Macaya, D., & Turgeon, C. T. (2009). Newborn screening for X-linked adrenoleukodystrophy (X-ALD): validation of a combined liquid chromatography-tandem mass spectrometric (LC-MS/MS) method. Molecular genetics and metabolism, 97(3), 212-220. https://doi.org/10.1016/j.ymgme.2009.03.010

Kilgore, M. B., Platis, D., Lim, T., Isenberg, S., Pickens, C. A., Cuthbert, C., & Petritis, K. (2023). Development of a universal second-tier newborn screening LC-MS/MS method for amino acids, lysophosphatidylcholines, and organic acids. Analytical chemistry, 95(6), 3187-3194. https://doi.org/10.1021/acs.analchem.2c03098

Klippel, C., Park, J., Sandin, S., Winstone, T. M., Chen, X., Orton, D., Singh, A., Hill, J. D., Shahbal, T. K., & Hamacher, E. (2025). Advancing newborn screening in Washington State: a novel multiplexed LC-MS/MS proteomic assay for Wilson disease and inborn errors of immunity. International journal of neonatal screening, 11(1), 6. https://doi.org/10.3390/ijns11010006

la Marca, G., Malvagia, S., Casetta, B., Pasquini, E., Donati, M. A., & Zammarchi, E. (2008). Progress in expanded newborn screening for metabolic conditions by LC-MS/MS in Tuscany: update on methods to reduce false tests. Journal of Inherited Metabolic Disease: Official Journal of the Society for the Study of Inborn Errors of Metabolism, 31, 395-404. https://doi.org/10.1007/s10545-008-0965-z

Lawrence, S. E., Marcadier, J., Auger, S., Bariciak, E., Chakraborty, P., Chambers, A., Eagle, S., Kowalski, M., Marr, A., & McIntosh, N. (2026). Improved performance of newborn screening for congenital adrenal hyperplasia using 21-deoxycortisol measurement. Journal of the Endocrine Society, 10(1), bvaf157. https://doi.org/10.1210/jendso/bvaf157

Mak, J., Peng, G., Le, A., Gandotra, N., Enns, G. M., Scharfe, C., & Cowan, T. M. (2023). Validation of a targeted metabolomics panel for improved second-tier newborn screening. Journal of inherited metabolic disease, 46(2), 194-205. https://doi.org/10.1002/jimd.12591

Matuszewski, B. K., Constanzer, M., & Chavez-Eng, C. (2003). Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC− MS/MS. Analytical chemistry, 75(13), 3019-3030. https://doi.org/10.1021/ac020361s

McHugh, D., Cameron, C. A., Abdenur, J. E., Abdulrahman, M., Adair, O., Al Nuaimi, S. A., Åhlman, H., Allen, J. J., Antonozzi, I., & Archer, S. (2011). Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. Genetics in Medicine, 13(3), 230-254. https://doi.org/10.1097/GIM.0b013e31820d5e67

Menkovic, I., Marchand, A.-S., Boutin, M., & Auray-Blais, C. (2019). Neonatal mass urine screening approach for early detection of mucopolysaccharidoses by UPLC-MS/MS. Diagnostics, 9(4), 195. https://doi.org/10.3390/diagnostics9040195

Millington, D., Kodo, N., Norwood, D., & Roe, C. (1990). Tandem mass spectrometry: a new method for acylcarnitine profiling with potential for neonatal screening for inborn errors of metabolism. Journal of inherited metabolic disease, 13(3), 321-324. https://doi.org/10.1007/BF01799385

Oboshi, W., Hirakiyama, A., Miura, M., Omachi, H., Tanaka, M., Seo, J.-H., Ohtake, A., Osawa, Y., Ishige, M., & Kosuga, M. (2026). Quantification of glycosaminoglycans in dried blood spots, and evaluation of its usefulness as a secondary newborn screening test for mucopolysaccharidoses. Biochemistry and Biophysics Reports, 46, 102569. https://doi.org/10.1016/j.bbrep.2026.102569

Oguni, T., Tomatsu, S., Tanaka, M., Orii, K., Fukao, T., Watanabe, J., Fukuda, S., Notsu, Y., Vu, D. C., & Can, T. B. N. (2020). Validation of liquid chromatography-tandem mass spectrometry-based 5-plex assay for mucopolysaccharidoses. International Journal of Molecular Sciences, 21(6), 2025. https://doi.org/10.3390/ijms21062025

Ombrone, D., Giocaliere, E., Forni, G., Malvagia,

S., & la Marca, G. (2016). Expanded newborn screening by mass spectrometry: new tests, future perspectives. Mass spectrometry reviews, 35(1), 71-84. https://doi.org/10.1002/mas.21463

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj, 372. https://doi.org/10.1136/bmj.n71

Pajares-García, S., González de Aledo-Castillo, J. M., Flores-Jiménez, J. E., Collado, T., Pérez, J., Paredes-Fuentes, A. J., Argudo-Ramírez, A., López-Galera, R. M., Prats, B., & García-Villoria, J. (2024). Analysis of a second-tier test panel in dried blood spot samples using liquid chromatography-tandem mass spectrometry in Catalonia's newborn screening programme. Clinical Chemistry and Laboratory Medicine (CCLM), 62(3), 493-505. https://doi.org/10.1515/cclm-2023-0216

Phoungphosop, J., Arpornsuwan, T., Jaresitthikunchai, J., Phaonakrop, N., Thaisakul, S., Thongngao, P., Thiplakorn, P., Roytrakul, S., & Krasao, P. (2026). Rapid peptide analysis in dried bloodspots to identify novel markers for newborn screening for congenital hypothyroidism. Scientific Reports, 16(1), 12955. https://doi.org/10.1038/s41598-026-42578-w

Pickens, C. A., Sternberg, M., Seeterlin, M., De Jesús, V. R., Morrissey, M., Manning, A., Bhakta, S., Held, P. K., Mei, J., & Cuthbert, C. (2020). Harmonizing newborn screening laboratory proficiency test results using the CDC NSQAP reference materials. International journal of neonatal screening, 6(3), 75. https://doi.org/10.3390/ijns6030075

Posern, C., Dreyer, B., Maier, S. L., Eichler, F., Gelb, M. H., Santer, R., Bley, A., & Murko, S. (2024). Quantification of N-acetyl-l-aspartate in dried blood spots: A simple and fast LC-MS/MS neonatal screening method for the diagnosis of Canavan disease. Molecular genetics and metabolism, 142(2), 108489. https://doi.org/10.1016/j.ymgme.2024.108489

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Systematic reviews, 10(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Saeed, Z., Mashkani, B., Alaei, A., Varasteh, A., & Keyfi, F. (2024). Determining the reference range of amino acids in healthy neonatal blood samples in northeast Iran using LC-MS/MS. Reports of Biochemistry & Molecular Biology, 13(1), 87. https://doi.org/10.61186/rbmb.13.1.87

Sidorina, A., Catesini, G., Deodato, F., Boenzi, S., Martinelli, D., Rizzo, C., & Dionisi-Vici, C. (2026). New multiplex LC-MS/MS method for lipid biomarker analysis of inherited neurodegenerative metabolic diseases. Journal of Lipid Research, 67(1). https://doi.org/10.1016/j.jlr.2025.100967

Stapleton, M., Kubaski, F., Mason, R. W., Shintaku, H., Kobayashi, H., Yamaguchi, S., Taketani, T., Suzuki, Y., Orii, K., & Orii, T. (2020). Newborn screening for mucopolysaccharidoses: Measurement of glycosaminoglycans by LC-MS/MS. Molecular Genetics and Metabolism Reports, 22, 100563. https://doi.org/10.1016/j.ymgmr.2019.100563

Therrell, B. L., Padilla, C. D., Loeber, J. G., Kneisser, I., Saadallah, A., Borrajo, G. J., & Adams, J. (2015). Current status of newborn screening worldwide: 2015. Seminars in perinatology, https://doi.org/10.1053/j.semperi.2015.03.002

Wang, C., Zhu, H., Zhang, W., Song, F., Liu, Z., & Liu, S. (2013). Second-tier test for quantification of underivatized amino acids in dry blood spot for metabolic diseases in newborn screening. Amino acids, 44(2), 661-671. https://doi.org/10.1007/s00726-012-1389-5

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., & Group*, Q.-. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine, 155(8), 529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-

Genomics and AI in Personalized Medicine
Volume 1, Issue 1
ISSN: 3071-4583

00009

Xie, Y., Peng, G., Tikhonova, I., Enns, G., Zhao, H., Cowan, T., & Scharfe, C. (2025). Improving newborn screening accuracy through genome sequencing, targeted metabolomics, and machine learning. BMC Medical Genomics, 18(1), 187. https://doi.org/10.1186/s12920-025-02261-x

Yusuf, C., Sontag, M. K., Miller, J., Kellar-Guenther, Y., McKasson, S., Shone, S., Singh, S., & Ojodu, J. (2019). Development of national newborn screening quality indicators in the United States. International journal of neonatal screening, 5(3), 34. https://doi.org/10.3390/ijns5030034