



# Phenylketonuria from Classical Phenotypes to Base Editing and Targeted Small Molecules: A Critical Narrative Review of Mechanistic Advance and Translational Uncertainty

**Stefan Bittmann<sup>a,b,++\*</sup>, Elisabeth Luchter<sup>a</sup>  
and Elena Moschüring-Alieva<sup>a</sup>**

<sup>a</sup> Department of Pediatrics, Ped Mind Institute, Department of Pediatrics, Hindenburgring 4, D-48599 Gronau, Germany.

<sup>b</sup> School of Medicine, Shangluo Vocational and Technical College, Shangluo, 726000, Shaanxi, China.

## *Authors' contributions*

*This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.*

## *Article Information*

DOI: <https://doi.org/10.9734/jammr/2026/v38i106221>

## **Open Peer Review History:**

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://pr.sdiarticle5.com/review-history/167058>

## *Review Article*

**Received: 02/08/2026**

**Accepted: 27/09/2026**

**Published: 29/09/2026**

## **Abstract**

Phenylalanine hydroxylase deficiency, historically described as phenylketonuria, occupies an unusual position in medicine. It was the first inherited metabolic disorder for which dietary intervention prevented intellectual disability, the first to be detected through population newborn screening, and it is now among the first for which *in vivo* genome correction has reached advanced preclinical validation. This review examines whether the conceptual architecture inherited from that history, above all the classification of patients into discrete phenotypic categories defined by pre-treatment blood phenylalanine concentration, remains adequate for a

<sup>++</sup> Honorary Professor in Pediatrics;

\*Corresponding author: E-mail: [stefanbittmann@gmx.de](mailto:stefanbittmann@gmx.de);

therapeutic landscape organised around residual enzyme function, variant-specific pharmacology and allele-specific correction. Literature was identified through structured searching of biomedical and multidisciplinary scholarly sources, supplemented by backward and forward citation tracking and by targeted searches for corrections and post-publication correspondence, with critical appraisal focused on design adequacy, endpoint validity and the distance between mechanistic demonstration and demonstrated clinical benefit. Four findings structure the synthesis. First, phenotypic classification predicts population-level severity but performs inconsistently at the level of the individual patient, and its predictive limits are inherited by every therapy selected on its basis. Second, the residual morbidity of early and continuously treated disease is real but modest in average magnitude, heterogeneous in expression and inconsistently attributable to phenylalanine exposure rather than to the restrictive diet itself. Third, pharmacological options have expanded substantially, and a recent head-to-head trial provides the first randomised comparison between cofactor-based agents, yet the evidence base remains dominated by biochemical surrogates over short horizons. Fourth, adenine base editing has achieved durable normalisation of blood phenylalanine in variant-humanised mouse models and improvement in brain biochemistry and motor phenotype, but no corrective editing approach has published peer-reviewed clinical outcomes, and the variant heterogeneity of the disorder poses a scalability problem that current allele-specific strategies do not resolve. The field's principal unresolved question is no longer whether phenylalanine can be lowered, but which outcomes justify which risks, in whom, and at what age.

**Keywords:** Phenylalanine hydroxylase deficiency; phenylketonuria; adenine base editing; sepiapterin; pegvaliase; genotype–phenotype correlation; tetrahydrobiopterin; lipid nanoparticles.

## 1. Introduction

Phenylalanine hydroxylase (PAH; EC 1.14.16.1) catalyses the tetrahydrobiopterin (BH4)-dependent para-hydroxylation of L-phenylalanine (Phe) to L-tyrosine, the committed and rate-limiting step of phenylalanine catabolism in mammalian liver. Biallelic pathogenic variation in *PAH* reduces or abolishes this flux, and the resulting accumulation of phenylalanine in blood and brain produces, when untreated, a syndrome of progressive intellectual disability, epilepsy, movement disorder and behavioural disturbance (van Spronsen *et al.*, 2021). The disorder is the most common autosomal recessive Mendelian phenotype of amino acid metabolism, with an estimated global prevalence of one in 23,930 live births and approximately 0.45 million affected individuals worldwide, although the range across populations is wide, extending from roughly one in 4,500 in Italy to one in 125,000 in Japan (Hillert *et al.*, 2020).

The therapeutic history of the condition is unusually well defined. Følling's identification of phenylpyruvic acid in the urine of affected siblings, Bickel's demonstration that a phenylalanine-restricted diet reversed biochemical and clinical deterioration, and Guthrie's bacterial inhibition assay together converted an untreatable cause of institutionalisation into a screened and managed chronic disease within three decades (Woolf & Adams, 2020). That achievement created the clinical vocabulary still in use. Because treatment decisions in the 1960s depended on a single measurable quantity, patients were stratified by pre-treatment or untreated blood phenylalanine concentration into classical phenylketonuria, milder phenylketonuria variants, and mild hyperphenylalaninaemia. The global distribution of these categories has been quantified in the largest genotype–phenotype dataset assembled to date, in which 62% of 16,092 affected individuals were classified as having classic phenylketonuria, 22% mild phenylketonuria and 16% mild hyperphenylalaninaemia, with systematic geographical gradients in both genotype and phenotype (Hillert *et al.*, 2020).

Three developments have placed that inherited framework under strain. The first is pharmacological. Sapropterin dihydrochloride, a synthetic formulation of the natural cofactor, acts in part as a pharmacological chaperone that stabilises mutant PAH protein, and its benefit is therefore concentrated in individuals retaining folding-competent enzyme (Muntau *et al.*, 2019; Pey *et al.*, 2008). Pegvaliase, a pegylated recombinant phenylalanine ammonia lyase derived from *Anabaena variabilis*, bypasses PAH entirely and is consequently indifferent to genotype, but requires immune tolerisation and carries a hypersensitivity burden (Harding *et al.*, 2018; Longo *et al.*, 2019). Sepiapterin, a naturally occurring BH4 precursor approved in 2025, reached a broader responder population than sapropterin in a phase 3 placebo-controlled trial that deliberately included individuals with classic disease and prior BH4 non-responders (Lamb, 2026; Muntau *et al.*, 2024). Each of these agents partitions patients differently, and none partitions them along the classical phenotype boundaries. The first revision of the European PKU guidelines, published in 2025, updated earlier recommendations and incorporated additional topics while acknowledging that important evidence gaps persist, reinforcing the need for careful evidence appraisal as therapeutic options expand (van Wegberg *et al.*, 2025).

The second development is genomic. The *PAH* locus is extraordinarily heterogeneous: 758 variants were catalogued in the international genotype dataset, most genotypes were compound heterozygous, and 55% of 3,659 distinct genotypes were observed in a single individual (Hillert *et al.*, 2020). Quantitative models that assign allelic phenotype values to individual variants and combine them predictively have improved population-level correlation but do not eliminate discordant individual cases (Garbade *et al.*, 2019). Contemporary single-centre series continue to report that a small number of variants account for over half of alleles while BH4 responsiveness distributes unevenly even within apparently similar genotypes (Karaca *et al.*, 2026).

Recent functional profiling across common PAH genotypes further indicates that residual enzyme activity and sapropterin responsiveness can be integrated to support genotype-informed treatment stratification (Gundorova *et al.*, 2025). Complementary work evaluating the 13C-phenylalanine breath test reflects a parallel effort to quantify *in vivo* residual PAH activity as an individual-level measure relevant to personalised treatment planning (Haitjema *et al.*, 2026).

The third development is technological, and it is the reason a new synthesis is warranted. Adenine base editing and prime editing permit direct correction of individual pathogenic substitutions in hepatocytes without double-strand break formation. Within four

years, investigators have demonstrated complete normalisation of blood phenylalanine in variant-humanised mice carrying the two most frequent pathogenic alleles, extended correction to further recurrent variants, and shown improvement in cerebral amino acid and neurotransmitter concentrations together with motor phenotype after a single lipid nanoparticle (LNP) dose (Brooks, Carrasco, *et al.*, 2023; Brooks, Whittaker, *et al.*, 2023; Gautier *et al.*, 2026; Quigley *et al.*, 2025). These results are mechanistically decisive in a way that earlier gene addition strategies were not. They also raise a problem that the classical framework cannot express: a therapy that corrects c.1222C>T is not a therapy for phenylketonuria, it is a therapy for one allele of it.

A recent global review of lifelong PKU management similarly emphasises that therapeutic development must be considered alongside persistent challenges in treatment access, adherence and implementation across settings (Longo *et al.*, 2025).

Existing reviews address parts of this terrain but leave the junction between them unexamined. Authoritative guidelines set management standards without adjudicating the conceptual adequacy of phenotypic classification (Smith *et al.*, 2025; van Spronsen *et al.*, 2017; van Wegberg *et al.*, 2017). Gene therapy reviews survey vector and editor platforms with limited engagement with the pharmacological alternatives against which editing will be judged (Martinez *et al.*, 2024; Yu & Jeong, 2025). Nutritional reviews treat dietary management largely in isolation from the pharmacological agents now reshaping dietary tolerance (Jones & Strehle, 2026). A recent commentary has raised the scalability problem explicitly but as a short opinion piece rather than an evidence synthesis (Nicosia, 2026). No recent critical review has asked whether the categories used to select patients for these interventions are fit for the purpose to which they are now being put.

## 1.1 Scope, Research Gap and Objectives

This review addresses phenylalanine hydroxylase deficiency in humans across the lifespan, together with the preclinical models used to develop therapies for it. Its central problem is the relationship between phenotypic and genotypic classification on one hand, and therapeutic selection, outcome definition and evidential adequacy on the other. Four objectives follow. The first is to evaluate how far classical phenotypic categories retain analytical utility once residual enzyme function, cofactor responsiveness and allele identity are available as alternative stratifiers. The second is to appraise the evidence for residual morbidity in early and continuously treated disease, distinguishing effects attributable to phenylalanine exposure from those attributable to the restrictive diet or to ascertainment. The third is to compare approved and investigational therapies on the strength and design of the evidence supporting them rather than on mechanistic appeal. The fourth is to assess how close corrective genome editing stands to clinical application, and what would have to be demonstrated for it to displace chronic pharmacological management.

The review deliberately excludes primary disorders of BH4 synthesis and regeneration, which present with hyperphenylalaninaemia but differ in pathophysiology, neurotransmitter involvement and management. It does not attempt a comprehensive account of newborn screening programme design, of medical food formulation chemistry, or of vector manufacturing. It does not quantify pooled treatment effects, and it does not rank therapies, because the available trials differ too greatly in population, comparator and endpoint to support such ranking. Its intended contribution is an integrated critical appraisal across a literature that is currently partitioned by discipline, and an explicit account of what the available evidence can and cannot support.

## 2. Methods for Literature Selection

### 2.1 Review Design and Rationale

A critical narrative design was chosen because the review question is conceptual and cross-disciplinary rather than a defined comparison of interventions against a common outcome. The relevant literature comprises randomised trials, observational cohorts, registry analyses, enzymological and structural studies, animal model experiments and consensus guidelines, which are not commensurable within a single effect estimate. A systematic review with meta-analysis would either have to restrict itself to one narrow comparison, forfeiting the integrative purpose, or pool heterogeneous designs in a way that would misrepresent the evidence. Narrative synthesis, however, carries recognised risks of selective citation and unexamined interpretive bias, and the reporting standards articulated for the quality assessment of narrative reviews were used as an internal framework to constrain them, specifically in relation to justification of importance, statement of aims, description of the literature search, referencing of claims, scientific reasoning and appropriate presentation of data (Baethge *et al.*, 2019). Transparency of sources and explicit appraisal were therefore treated as obligations of the design, not as features waived by it.

### 2.2 Information Sources

Searching was conducted in PubMed, Europe PMC, Crossref, the Directory of Open Access Journals, Semantic Scholar and Google Scholar. Crossref was additionally used as the primary registry for bibliographic verification. Searching of these sources was supplemented by backward citation searching of reference lists in recent reviews, consensus statements and trial reports; by forward citation searching of pivotal trials and preclinical proof-of-concept studies; by related-article retrieval; and by targeted searching for corrections, errata, editorial expressions of concern and post-publication correspondence relating to any guideline or trial cited. That last step identified a published correction and a substantive letter of correspondence relating to the 2023 evidence-based clinical guideline, both of which are treated as part of the evidence base rather than as incidental records (Burgard *et al.*, 2026; Smith *et al.*, 2025).

### 2.3 Search Strategy and Search Period

Search concepts were constructed around five axes: the disorder and enzyme, addressed through terms including phenylketonuria, hyperphenylalaninaemia, phenylalanine hydroxylase and *PAH*; classification and mechanism, addressed through genotype–

phenotype correlation, residual activity, protein misfolding and tetrahydrobiopterin responsiveness; established and emerging interventions, addressed through dietary restriction, protein substitute, glycomacropeptide, large neutral amino acids, sapropterin, sepiapterin, pegvaliase, phenylalanine ammonia lyase, gene therapy, adeno-associated virus, messenger RNA, base editing and prime editing; outcomes, addressed through blood phenylalanine concentration, dietary phenylalanine tolerance, cognitive and executive function, white matter integrity, bone mineral density, quality of life and adverse events; and cross-cutting constraints, addressed through off-target editing, durability, delivery, equity of access, cost-effectiveness and implementation. Concepts were combined with Boolean operators and adapted to the syntax of each source, using field-restricted forms where supported. A representative strategy took the form: ("phenylketonuria" OR "phenylalanine hydroxylase deficiency" OR "hyperphenylalaninaemia") AND ("base editing" OR "prime editing" OR "gene therapy" OR "adeno-associated virus" OR "lipid nanoparticle") AND ("correction" OR "editing efficiency" OR "off-target" OR "durability"). The search period extended from 1 January 2000 to 22 July 2026. Foundational publications predating this window were retained where historically or conceptually necessary. Search concepts and representative strings are summarised in Table 1.

**Table 1. Search concepts, representative Boolean strings and retrieval limitations**

| Concept axis                 | Representative terms and strings                                                                                                                                                                | Analytical purpose                                      | Principal retrieval limitation                                                     | Supporting references                                                                             |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Disorder and enzyme          | ("phenylketonuria" OR "phenylalanine hydroxylase deficiency" OR "hyperphenylalaninaemia" OR "PAH deficiency")                                                                                   | Define the population and exclude primary BH4 disorders | Inconsistent use of "phenylketonuria" and "PAH deficiency" as synonyms across eras | Hillert <i>et al.</i> (2020); van Spronsen <i>et al.</i> (2021)                                   |
| Classification and mechanism | ("genotype-phenotype" OR "allelic phenotype value" OR "residual activity" OR "protein misfolding" OR "BH4 responsiveness")                                                                      | Assess the basis and limits of phenotypic categories    | Functional annotation reported heterogeneously across in vitro systems             | Garbade <i>et al.</i> (2019); Karaca <i>et al.</i> (2026); Fitzpatrick (2023)                     |
| Interventions                | ("sapropterin" OR "sepiapterin" OR "pegvaliase" OR "glycomacropeptide" OR "large neutral amino acids" OR "base editing" OR "prime editing" OR "adeno-associated virus" OR "lipid nanoparticle") | Compare modalities on evidential strength               | Dietary literature dispersed across nutrition and metabolic indexing               | Muntau <i>et al.</i> (2024); Harding <i>et al.</i> (2024); Brooks, Carrasco, <i>et al.</i> (2023) |
| Outcomes                     | ("blood phenylalanine" OR "dietary Phe tolerance" OR "executive function" OR "white matter" OR "bone mineral density" OR "quality of life" OR "off-target")                                     | Distinguish surrogate from clinical endpoints           | Cognitive outcomes reported through non-comparable instruments                     | Romani <i>et al.</i> (2022, 2025); Whitehall <i>et al.</i> (2024); Kalter <i>et al.</i> (2025)    |
| Cross-cutting constraints    | ("scalability" OR "durability" OR "access" OR "cost-effectiveness" OR "newborn screening" AND "coverage")                                                                                       | Situate therapies within health-system reality          | Economic and access evidence concentrated in high-income settings                  | Muntau <i>et al.</i> (2026); Trampuž <i>et al.</i> (2025); Nicosia (2026)                         |

*Note.* BH4 = tetrahydrobiopterin; PAH = phenylalanine hydroxylase; Phe = phenylalanine. Terms are shown in the form used for concept construction; field tags and truncation were adapted to each source

## 2.4 Eligibility Criteria

Peer-reviewed original research, systematic reviews, meta-analyses, consensus statements and clinical practice guidelines were eligible. Included human studies addressed individuals of any age with phenylalanine hydroxylase deficiency; included preclinical studies used cellular or animal models of the disorder, with priority given to variant-humanised models in which the human pathogenic allele is present. Eligible outcomes comprised biochemical control, dietary phenylalanine tolerance, neurocognitive and neuropsychiatric measures, neuroimaging indices, somatic comorbidity, patient-reported outcomes, editing efficiency, off-target profiles, durability and safety. Publications in English were eligible. Excluded were conference abstracts without full peer-reviewed reports, preprints where peer-reviewed evidence on the same question existed, commercial and promotional material, patents used as evidence of effectiveness, and any record whose bibliographic identity could not be confirmed. Studies of primary BH4 synthesis and regeneration disorders were excluded except where cited to explain the diagnostic boundary.

## 2.5 Screening, Duplicate Handling and Study Selection

Titles and abstracts retrieved from each source were assessed against the eligibility criteria, and records passing that stage were examined in full text where accessible. Duplicates arising from overlapping coverage between sources were identified by matching

Digital Object Identifiers and, where absent, by matching first author, title and year. Where a record appeared as both a preprint and a peer-reviewed article, the peer-reviewed version was retained. Where interim and final reports of the same trial programme were available, both were retained when they addressed different questions and the later report was preferred when they addressed the same question. Records whose bibliographic details could not be verified against an authoritative registry were not used. Because selection was purposive and iterative rather than exhaustive, no screening counts are reported and no flow diagram is presented; representing a narrative selection process in those terms would misstate what was done.

## 2.6 Critical Appraisal and Evidence Synthesis

Each included study was appraised against criteria appropriate to its design. For randomised trials, attention was given to allocation and blinding, the adequacy of the comparator, whether the primary endpoint was a clinical outcome or a biochemical surrogate, trial duration relative to the chronicity of the disorder, handling of responder enrichment, and sponsor involvement. For observational and registry studies, attention was given to representativeness, confounding by indication, completeness of follow-up and the distinction between association and effect. For meta-analyses, attention was given to the heterogeneity of the constituent studies and to whether pooled estimates were interpretable given that heterogeneity. For preclinical work, attention was given to whether the model carried the human pathogenic variant, whether efficacy was demonstrated in the relevant target tissue, the completeness of off-target characterisation, and the duration over which effect was sustained. Formal risk-of-bias instruments were not applied, because doing so selectively across designs would convey a false impression of systematic appraisal. Themes were developed inductively around mechanisms, decision points and points of disagreement rather than around study types, and conflicting findings were retained and examined rather than reconciled.

Table 1 records the concepts and representative strings used, together with the analytical purpose each axis served and the principal limitation encountered in retrieving it. The table is presented to make the search reproducible in structure rather than to claim exhaustive coverage; the literature on dietary management in particular is published across nutrition, metabolic medicine and dietetics with inconsistent indexing, and the table records this as a retrieval limitation rather than concealing it.

## 3. From Classical Phenotypes to a Graded Metabolic Spectrum

### 3.1 The Historical Construction of the Classical Phenotype

The tripartite classification of phenylalanine hydroxylase deficiency was an artefact of what could be measured when the disorder became treatable. Untreated blood phenylalanine concentration was the only quantity available in the 1960s that correlated with the risk of intellectual disability, and it was accordingly made to carry the entire diagnostic and prognostic load (Woolf & Adams, 2020). The categories worked because they were used for a single purpose: to decide who required dietary restriction and how severe that restriction should be. Their apparent biological solidity came later, through repetition rather than validation.

That origin matters because the classification embeds two assumptions that subsequent evidence has qualified. The first is that untreated phenylalanine concentration is a stable individual property. It is not: phenylalanine concentration reflects dietary intake, catabolic state, growth and the assay matrix, and international consensus on evaluating cofactor responsiveness has had to specify standardised loading conditions precisely because uncontrolled measurement conditions generate misclassification (Muntau *et al.*, 2019). The second assumption is that the categories are separated by natural discontinuities. The distribution of residual PAH activity across pathogenic genotypes is better described as continuous, and the boundaries between classical phenylketonuria, mild phenylketonuria and mild hyperphenylalaninaemia fall at conventional rather than biological thresholds (Hillert *et al.*, 2020). The transition from a phenotype-led to a variant-led understanding of the disorder has been traced in detail, and its central observation is that the accumulation of molecular data progressively dissolved the categorical picture that preceded it (Blau, 2016).

The consequence is not that the categories are useless. They retain descriptive value at population scale, and the global phenotype distribution reported across 51 countries demonstrates real and geographically structured variation in disease severity that any account of the disorder must accommodate (Hillert *et al.*, 2020). The consequence is that they carry more explanatory weight in contemporary use than their construction supports, and that therapies selected on their basis inherit that weakness.

### 3.2 Enzymology and the Structural Basis of Residual Function

Understanding why the phenotypic spectrum is continuous requires attention to the enzyme itself. PAH is a homotetrameric, non-haem iron-dependent aromatic amino acid hydroxylase in which each subunit comprises a regulatory domain, a catalytic domain and an oligomerisation domain. Its activity is governed by an allosteric mechanism in which phenylalanine binding to the regulatory domain stabilises an activated conformation, while BH4 binding favours a less active resting state, so that enzyme activity responds to substrate availability with sigmoidal rather than hyperbolic kinetics (Fitzpatrick, 2023). This architecture explains several features of the clinical disorder that a simple loss-of-function model does not. Because activation depends on conformational transitions distributed across domains, variants remote from the catalytic site can impair function; because the tetramer assembles from subunits, compound heterozygosity can generate hybrid tetramers whose activity is not the arithmetic mean of the two alleles; and because many pathogenic variants destabilise the protein rather than abolishing catalysis, residual activity is a function of cellular protein quality control as much as of intrinsic enzymatic competence.

That last point is the mechanistic basis of pharmacological chaperone therapy. BH4 binding stabilises the folded state of destabilised mutant PAH, increasing the steady-state quantity of functional enzyme rather than the specific activity of each molecule, and chemical chaperones acting by the same principle but binding at distinct sites were identified as candidate therapeutics and shown to

correct the phenotype in mouse models (Pey *et al.*, 2008; Santos-Sierra *et al.*, 2012). The therapeutic implication is direct: cofactor responsiveness is predicted by folding competence, not by phenotype category, which is why cofactor response distributes across classical boundaries rather than respecting them.

The evidence here is strong in its mechanistic core and weaker at its margins. Structural and kinetic characterisation of PAH regulation rests on convergent crystallographic, scattering and enzymological work and is not seriously contested (Fitzpatrick, 2023). The extrapolation from *in vitro* stabilisation to *in vivo* clinical response is less secure, because cellular chaperone networks, hepatic protein turnover and substrate flux are not reproduced in purified systems. The chemical chaperone work established proof of principle over a decade ago (Santos-Sierra *et al.*, 2012), yet no compound from that line has entered routine clinical use, which is itself informative about the difficulty of the translational step rather than about the validity of the mechanism.

### 3.3 Genotype–Phenotype Correlation and the Limits of Prediction

If phenotypic categories are conventional, genotype offers the obvious alternative stratifier, and considerable effort has been directed to making it quantitative. The allelic phenotype value approach assigns each variant a numerical severity derived from observed phenotypes in individuals carrying it, then combines the two allelic values into a genotype-level prediction, substantially improving on categorical assignment of variants as severe or mild (Garbade *et al.*, 2019). Applied at scale, such methods explain a large proportion of phenotypic variance and permit predictions of expected metabolic phenotype at diagnosis.

The limits are equally clear and are not adequately acknowledged in clinical discourse. Prediction is weakest precisely where it would be most useful: in compound heterozygotes, which constitute 73% of genotypes, and in rare genotypes, of which 55% were observed only once in the largest international dataset (Hillert *et al.*, 2020). A variant seen once cannot be assigned an empirically derived severity value, and a genotype seen once cannot be validated. Contemporary cohort studies reinforce the point from the other direction. In a series of 171 individuals ascertained through newborn screening, 63 distinct *PAH* variants were identified, five variants accounted for more than half of the cohort, and among 100 individuals tested, 41% responded to BH4 loading, with responsiveness concentrated in but not confined to particular variants (Karaca *et al.*, 2026). The structure of the data is consistently the same: a small number of common alleles that behave predictably, surrounded by a long tail of rare alleles that do not.

Two further considerations complicate the interpretation. First, genotype–phenotype series are subject to ascertainment effects. Cohorts assembled through newborn screening in one country reflect that population's allele frequencies, and generalisation across populations is constrained by the geographical gradients in genotype distribution that have been documented directly (Hillert *et al.*, 2020). Second, the outcome being predicted is usually untreated or pre-treatment phenylalanine concentration, which returns the analysis to the same conventional categories whose adequacy is in question. Genotype-based prediction of a conventionally defined phenotype cannot establish that the phenotype is the right target of prediction. Prediction of cofactor responsiveness, dietary tolerance or long-term neurocognitive outcome would be more informative, and such models remain comparatively underdeveloped. The defensible position is therefore intermediate. Genotype adds real information beyond phenotype, particularly for common alleles and particularly for anticipating cofactor responsiveness, and it is reasonable that genetic testing at birth is now strongly recommended to confirm diagnosis and guide therapy (Smith *et al.*, 2025). It does not yet support individual-level prediction robust enough to determine therapy without biochemical confirmation, and the international best-practice framework for evaluating sapropterin responsiveness by standardised loading exists because genotype alone is insufficient (Muntau *et al.*, 2019).

Table 2 sets out the principal classification frameworks currently in use, the decision each was designed to support, and the specific analytical weakness each carries. The comparison makes visible a pattern that is easily missed when the frameworks are discussed separately: each was constructed to answer a question narrower than the question it is now used to answer, and the resulting mismatch is a recurring source of clinical uncertainty. The table also indicates where the frameworks disagree, since a patient may be classified as having classic disease by untreated phenylalanine concentration while carrying a genotype associated with partial cofactor responsiveness, and the therapeutic implications of the two classifications diverge.

**Table 2. Classification frameworks in phenylalanine hydroxylase deficiency, their intended decision support and analytical limitations**

| Framework                                                                        | Basis of classification                              | Decision originally supported                           | Principal analytical limitation                                                                                                   | Supporting references                                                           |
|----------------------------------------------------------------------------------|------------------------------------------------------|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Metabolic phenotype (classic, mild phenylketonuria, mild hyperphenylalaninaemia) | Untreated or pre-treatment blood Phe concentration   | Whether and how strictly to restrict dietary Phe        | Thresholds are conventional, not biological; measurement is condition-dependent; poor individual-level resolution near boundaries | Hillert <i>et al.</i> (2020); Muntau <i>et al.</i> (2019); Woolf & Adams (2020) |
| Variant-level functional annotation                                              | In vitro residual PAH activity and protein stability | Interpretation of novel variants; mechanistic inference | Assay systems are heterogeneous; hybrid tetramers in compound heterozygotes are not modelled                                      | Fitzpatrick (2023); Pey <i>et al.</i> (2008)                                    |

| Framework                                             | Basis of classification                                         | Decision originally supported                           | Principal analytical limitation                                                                                 | Supporting references                                                                 |
|-------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Allelic phenotype value and genotype-based prediction | Empirically derived severity values combined across two alleles | Prediction of expected metabolic phenotype at diagnosis | Undefined for rare alleles; 73% of genotypes are compound heterozygous and 55% of distinct genotypes occur once | Garbade <i>et al.</i> (2019); Hillert <i>et al.</i> (2020)                            |
| Cofactor responsiveness                               | Standardised BH4 or sapropterin loading response                | Selection for cofactor-based pharmacotherapy            | Response definitions differ across protocols and trials; responsiveness crosses phenotype categories            | Karaca <i>et al.</i> (2026); Muntau <i>et al.</i> (2019)                              |
| Allele identity for corrective editing                | Presence of a specific targetable pathogenic variant            | Eligibility for allele-specific genome editing          | Applies only to individuals carrying the targeted variant; excludes most rare alleles entirely                  | Brooks, Whittaker, <i>et al.</i> (2023); Nicosia (2026); Quigley <i>et al.</i> (2025) |

Note. BH4 = tetrahydrobiopterin; PAH = phenylalanine hydroxylase; Phe = phenylalanine. Frameworks are not mutually exclusive and are frequently applied simultaneously to the same individual, with divergent therapeutic implications

## 4. Residual Morbidity in Early and Continuously Treated Disease

### 4.1 Cognitive Outcome: Real, Modest and Heterogeneous

The justification for intensifying treatment, and ultimately for accepting the risks of genome editing, rests on the claim that early and continuously treated patients retain clinically meaningful impairment. That claim is supported, but its magnitude and structure are frequently overstated in advocacy for new therapies and understated in arguments for dietary relaxation.

The most rigorous synthesis available pooled 26 groups of early-treated adults, comprising 757 individuals contributing 220 measures across 19 cognitive functions, and found a weighted effect size of  $-0.44$  across all functions and  $-0.60$  for measured intelligence relative to matched controls (Romani *et al.*, 2022). Impairment was not uniform: effect sizes ranged from  $-1.02$  to  $-0.18$ , with the largest deficits in reasoning, visual-spatial attention speed, sustained attention, visuo-motor control and flexibility, and with consistently larger effects on speed than on accuracy measures and on visuospatial than on verbal material. A subsequent extension comparing 29 adult and 21 child groups, totalling 904 and 460 participants respectively, found a comparable overall impairment of approximately half a standard deviation in both age bands, but a different internal structure: children showed relatively homogeneous impairment across functions, whereas adults showed a wider spread and a marked dissociation between impaired speed and preserved accuracy (Romani *et al.*, 2025).

The interpretive implications deserve emphasis because they are often lost. An average deficit of half a standard deviation is a population-level shift, not a description of any individual, and a substantial proportion of treated adults perform within the normal range. The dissociation between speed and accuracy suggests compensation rather than static damage, and the authors' interpretation that adults draw on relatively preserved learning and crystallised ability to sustain accurate performance despite slowed processing is consistent with the observed pattern (Romani *et al.*, 2025). Both analyses also found that blood phenylalanine concentration modulated effect size in several conditions, with processing speed more strongly related to childhood than to concurrent adult concentrations, which supports the importance of early control while weakening the inference that adult concentrations drive adult cognitive status.

Against this, a double-blind randomised crossover trial in 30 adults with classical disease found that a four-week increase in phenylalanine intake, simulating unrestricted consumption, was non-inferior to placebo for working memory accuracy in both intention-to-treat and per-protocol analyses, with no significant differences in working memory reaction time, manual dexterity, mood or depression (Trepp *et al.*, 2024). This result is frequently cited in support of dietary relaxation in adults and is equally frequently dismissed. Neither response is adequate. The trial was well designed for its question and its negative finding is informative, but four weeks is a very short exposure relative to a lifelong disorder, the sample was small and single-site, and accuracy measures were precisely the domain that the meta-analytic evidence identifies as relatively preserved. The trial demonstrates that short-term phenylalanine elevation does not acutely degrade accuracy-based working memory in adults; it does not demonstrate that sustained elevation is without consequence, and it was not designed to.

### 4.2 Neuroimaging, Reversibility and the Mechanistic Gap

Structural neuroimaging has been used to bridge the gap between biochemistry and cognition, with white matter abnormality the most consistent finding. The question of clinical importance is whether such abnormality is fixed or responsive to improved control. A longitudinal study of adults undergoing intensified dietary management reported improvement in indices of white matter structure, indicating that at least some component of the imaging phenotype is dynamic rather than a permanent record of childhood exposure (Kang *et al.*, 2026). That finding is mechanistically encouraging and clinically suggestive, but its evidential weight is

limited by sample size and by the absence of a randomised comparator, and improvement in diffusion indices is not itself a patient-relevant outcome.

Mechanistic accounts of how phenylalanine damages the developing and mature brain remain plural rather than settled. Competition for large neutral amino acid transport at the blood–brain barrier, consequent depletion of cerebral tyrosine and tryptophan and impaired monoamine neurotransmitter synthesis, direct effects on myelin and oligodendrocyte function, and oxidative and mitochondrial perturbation have all been proposed and are not mutually exclusive (Nulmans *et al.*, 2025). Preclinical editing work provides some of the most direct evidence available, since correction of hepatic PAH in variant-humanised mice reduced cortical as well as plasma phenylalanine and increased cerebral amino acid and neurotransmitter concentrations, with concomitant improvement in abnormal motor phenotypes (Gautier *et al.*, 2026). That this occurred after a liver-directed intervention is important: it indicates that the cerebral phenotype in this model is driven by peripheral phenylalanine exposure and is at least partially reversible after the neonatal period, which is a stronger claim than correlation between blood concentration and imaging indices in humans can support.

Adults with the disorder also present with a distinct and under-recognised spectrum of neurological complications, including movement disorders and features requiring differentiation from other neurodegenerative and neuroinflammatory conditions, which has prompted explicit guidance on differential diagnosis (Merkel *et al.*, 2023). The existence of this spectrum argues against treating cognitive test scores as the sole relevant neurological outcome.

### 4.3 Somatic Comorbidity and the Attribution Problem

Beyond the brain, adults with the disorder experience excess somatic morbidity, but here the attribution problem is acute, because the exposure of interest and the treatment for it are confounded. A systematic review screening 1,185 records and including 51 studies documented somatic comorbidities across multiple organ systems in adults, with the central interpretive difficulty stated explicitly: high blood phenylalanine and chronic restriction of phenylalanine and intact protein are both candidate causes, and the available study designs rarely separate them (Whitehall *et al.*, 2024). A focused meta-analysis of bone mineral density drawing on 4,097 individuals across 10 studies of at least acceptable quality reached a similar impasse, with reduced bone mineral density plausibly attributable either to the metabolic consequences of the disorder or to the nutritional composition of long-term dietary management (Rocha *et al.*, 2024).

Large-scale claims data add epidemiological breadth without resolving causality. In a matched-cohort analysis of 5,729 pairs aged 12 years or older and 2,648 pairs under 12 years, individuals with the disorder showed significantly higher prevalence of neuropsychiatric and cognitive comorbidities than matched controls, with prevalence differences of 5.2% for intellectual or developmental disability, 4.9% for fatigue and malaise and 4.2% for anxiety, together with 39% more hospitalisations, 18% more emergency department visits and 25% more outpatient visits (Longo *et al.*, 2026). Such data establish that the treated condition carries a real and measurable burden on health systems and on patients. They cannot establish which component of that burden is attributable to phenylalanine exposure, which to the diet, which to the psychosocial consequences of lifelong chronic disease management, and which to detection bias arising from more frequent clinical contact.

Patient-reported outcomes have been mapped in a scoping review of quality of life in patients and caregivers, which found a literature limited by instrument heterogeneity and by the scarcity of disease-specific validated measures (Remor *et al.*, 2024). This is a substantive methodological obstacle rather than a minor one. If the argument for advancing to curative intervention rests on residual burden, and if the burden is measured with instruments that are neither disease-specific nor consistent between studies, then the comparator against which any new therapy will be judged is poorly characterised.

The honest summary is that residual morbidity in treated disease is demonstrated but imprecisely apportioned. Cognitive effects are modest on average, structured rather than global, and related more securely to childhood than to concurrent exposure. Somatic effects are documented but confounded between disease and treatment. Quality-of-life effects are reported but inadequately measured. Each of these limitations weakens the inferential chain that runs from residual morbidity to the justification for higher-risk therapies, and none of them is resolved by the volume of evidence currently available.

## 5. Dietary Management: Effective, Burdensome and Evidentially Thin

### 5.1 The Protein Substitute and Its Alternatives

Phenylalanine-restricted diet combined with a phenylalanine-free protein substitute remains the foundation of management and the only intervention with unambiguous evidence of preventing severe intellectual disability. Its evidential paradox is that this foundational status rests on historical cohort comparison rather than on randomised evidence, because withholding treatment became unethical before randomised trials were conducted. Everything subsequently tested has therefore been tested against a standard that was never itself randomised, and contemporary guidelines codify dietary practice on the basis of consensus and observational data rather than trial evidence (Smith *et al.*, 2025; van Wegberg *et al.*, 2017).

Within that constraint, the most substantial trial evidence concerns the composition of the protein substitute. Casein glycomacropeptide is an intact peptide naturally low in phenylalanine, proposed as more palatable and physiologically advantageous than free L-amino acid formulations. A two-stage randomised crossover trial in 30 early-treated participants aged 15 to 49 years, of whom 20 had classical disease, compared glycomacropeptide and amino acid medical foods providing equal protein equivalents over three-week periods (Ney *et al.*, 2016). Intake frequency was higher with glycomacropeptide, participants rated it more

acceptable and reported improved gastrointestinal symptoms and reduced hunger, and analysis of covariance showed no significant mean increase in plasma phenylalanine despite a significantly higher phenylalanine intake from the glycomacropeptide product, whereas the amino acid product significantly decreased plasma phenylalanine. The honest reading is that glycomacropeptide improved acceptability and tolerability at the cost of slightly poorer biochemical control, a trade-off rather than a superiority.

A randomised crossover trial in children reached a similar structure of result over a longer exposure. Twelve children aged four to nine years received each product as the sole protein substitute for 12 weeks, separated by a four-week washout, with gut health, oxidative stress, renal function, inflammatory markers and plasma amino acids assessed (Pinto *et al.*, 2024). Glycomacropeptide significantly improved stomach pain, heartburn and reflux, and wind and bloating. The sample of 12 is very small, and although the crossover design increases efficiency, the study is not powered for biochemical or safety endpoints and should be read as evidence about symptoms rather than about metabolic adequacy.

The question of nutritional equivalence has been addressed more directly using stable isotope methodology. In six adults studied with the indicator amino acid oxidation technique across seven L-amino acid intakes and four glycomacropeptide intakes, protein requirements and the bioavailability of the glycomacropeptide product were determined against an isotopically defined indicator of protein synthesis (Turki *et al.*, 2025). This is a methodologically strong approach to a question that palatability trials cannot answer, but the sample of six limits inference, and the design measures short-term protein synthesis rather than long-term nutritional outcome.

Practice has nonetheless moved. A survey of health professionals across multiple countries documented that glycomacropeptide substitutes are now widely used internationally, with considerable variation in how and to whom they are prescribed (Evans *et al.*, 2026). The gap between the modest trial base and the breadth of adoption is a recurrent feature of nutritional management in this disorder, and it is not adequately acknowledged in guideline documents that present dietary recommendations with a confidence their evidential base does not warrant.

## 5.2 Large Neutral Amino Acids

Supplementation with large neutral amino acids targets a different mechanism: competition with phenylalanine for the L-type amino acid transporter at the blood–brain barrier and in the intestine, intended to reduce cerebral phenylalanine exposure without necessarily lowering plasma concentration (Rocha & Martel, 2009). The mechanistic rationale is sound and the approach has obvious appeal for adults unable or unwilling to maintain strict restriction.

The clinical evidence, however, is weak in ways that are seldom stated plainly. An open-label study of 10 adults with poor metabolic control treated for 12 months found no change in plasma phenylalanine, a significant improvement in tyrosine concentration, and improvement on psychometric testing (Scala *et al.*, 2020). With 10 participants, no control group, no blinding and repeated cognitive testing over a year, practice effects alone provide a sufficient alternative explanation for the psychometric findings. A separate study of a slow-release preparation reported improved treatment adherence in adults, an outcome that is plausible and clinically relevant but that was not assessed under blinded conditions (Burlina *et al.*, 2020). Notably, the absence of plasma phenylalanine reduction is consistent with the proposed mechanism rather than evidence against it, which is precisely why demonstration of benefit requires cerebral or cognitive endpoints measured under controlled conditions. Those studies have not been performed at adequate scale. The current position is that large neutral amino acid supplementation is mechanistically coherent, possibly useful as an adjunct in adults with poor control, and not supported by evidence that would justify it as a substitute for phenylalanine restriction.

## 5.3 Adherence, Transition and the Adult Treatment Gap

Dietary management fails not primarily because it is ineffective but because it is difficult to sustain. Structured transition from paediatric to adult services has been implemented and evaluated, with reported improvement in engagement over two years of follow-up with a dedicated adult team (Peres *et al.*, 2021). Such single-centre service evaluations are valuable as demonstrations of feasibility but cannot be generalised without replication, and they do not address populations lost before transition occurs.

The policy question underlying adherence has become sharper rather than clearer. The 2023 evidence-based guideline strongly recommends lifelong maintenance of blood phenylalanine at or below 360  $\mu\text{mol/L}$  on the grounds that individuals with lifelong concentrations at or below that threshold have better intellectual outcomes, that treatment should be lifelong for those with untreated concentrations above 360  $\mu\text{mol/L}$ , and that concentrations at or below 360  $\mu\text{mol/L}$  before conception are strongly recommended to prevent adverse pregnancy outcomes (Smith *et al.*, 2025). This represents a strengthening of recommendations that the preceding guideline had made without strong evidential support. The strengthening has been contested. Published correspondence from a group of European metabolic clinicians disputes aspects of the evidence appraisal and its clinical implications (Burgard *et al.*, 2026), and a separate multinational group has argued that treatment goals for adults with early-treated disease should be determined through evidence-based shared decision making between patients and clinical teams rather than by uniform target setting (Langeveld *et al.*, 2026). A published correction to the guideline further indicates that its content required post-publication amendment.

This disagreement is substantive and should not be smoothed over. The guideline's position is defensible on the evidence it appraises, particularly for childhood and for preconception management, where the case is strong. The objection is also defensible, particularly for adults with well-controlled childhood exposure, where the meta-analytic evidence indicates that concurrent adult concentrations are less strongly related to processing speed than childhood concentrations were (Romani *et al.*, 2025) and where a

randomised challenge found no short-term decrement in accuracy-based working memory (Trepp *et al.*, 2024). What the field lacks is the long-term randomised or quasi-experimental evidence in adults that would settle the question, and neither position can claim that evidence.

Maternal management is the one domain in which the case for strict control is not seriously disputed, given the teratogenic consequences of elevated maternal phenylalanine. Even here, implementation is imperfect: factors determining good metabolic control throughout pregnancy have been examined in detail (Rohde *et al.*, 2021), and a recent multinational analysis of 102 pregnancies documented the extent to which target concentrations are achieved in European practice (Pinto *et al.*, 2026). The gap between recommendation and achievement in the least contested area of management is informative about the realism of expectations elsewhere.

## 6. Targeted Small Molecules and Enzyme Substitution

### 6.1 Sapropterin: Chaperone Pharmacology and the Responder Problem

Sapropterin dihydrochloride was the first pharmacological therapy for the disorder and remains the reference against which newer cofactor-based agents are judged. Its mechanism, stabilisation of folding-compromised mutant PAH with a consequent increase in functional enzyme, follows directly from the enzymology and from the chaperone work that preceded clinical development (Pey *et al.*, 2008; Santos-Sierra *et al.*, 2012).

The clinical evidence is substantial in volume and dominated by registry and observational designs. The KAMPER registry enrolled 576 participants with PAH deficiency across nine European countries and 69 sites, with follow-up extending over more than a decade. Adverse events occurred in 401 participants (69.6%), totalling 1,960 events, of which 61 events in 42 participants were serious and two were considered sapropterin-related by the investigator; mean actual dietary phenylalanine intake increased from 957 mg/day at baseline to a maximum of 1,959 mg/day across the study period (Feillet *et al.*, 2025). The safety signal is reassuring and the tolerance gain is clinically meaningful, since increased dietary phenylalanine tolerance translates directly into dietary liberalisation. The design limits are equally clear: a registry of treated patients has no untreated comparator, enrolment favours responders by construction, and attribution of dietary tolerance change to drug rather than to concurrent changes in clinical management cannot be secured. Within the same registry programme, analyses of sapropterin use before and during pregnancy provide reassurance on a question that randomised designs cannot address (Feillet *et al.*, 2024).

The persistent difficulty is definition of response. International best-practice guidance exists precisely because loading protocols, response thresholds and measurement conditions varied enough to generate inconsistent classification across centres (Muntau *et al.*, 2019). Response thresholds of 20% or 30% reduction in blood phenylalanine are conventions, and a patient classified as a non-responder under one protocol may be a responder under another. Any comparison of cofactor agents across trials is therefore partly a comparison of response definitions.

### 6.2 Sepiapterin: Broader Response and the Question of Comparator

Sepiapterin, a natural precursor that is converted intracellularly to BH4, reached regulatory approval in 2025 (Lamb, 2026). Its evidential base is the strongest of any pharmacological agent in this disorder, and it repays close reading. Clinical development proceeded from a phase 2 randomised, multicentre, three-period crossover, open-label, active-controlled study in an all-comers population, which established dose-related phenylalanine lowering and supported progression to a definitive trial (Bratkovic *et al.*, 2022).

APHENITY was a phase 3, randomised, double-blind, placebo-controlled trial conducted at 34 sites in 13 countries, enrolling individuals of all ages with blood phenylalanine of 360  $\mu\text{mol/L}$  or higher and excluding primary BH4 deficiencies (Muntau *et al.*, 2024). The design comprised a 14-day open-label response assessment followed by randomisation of responders to six weeks of sepiapterin with forced dose escalation, or placebo. Of 156 participants assessed in part 1, 114 (73%) were sepiapterin-responsive by the prespecified criterion of at least 15% reduction in blood phenylalanine, and 98 participants entered the randomised primary analysis set, which required at least 30% reduction in part 1. A significant reduction in blood phenylalanine was demonstrated for sepiapterin relative to placebo, and the trial's design permitted assessment across a deliberately broad population including individuals with classic disease, individuals receiving sapropterin at entry, and prior BH4 non-responders.

Three critical observations follow. First, the enrichment structure is consequential. Randomisation occurred after responder selection, so the randomised estimate describes the effect in responders, not in the population presenting for treatment; the 73% response rate in part 1 is the more relevant figure for clinical planning and it was measured without a control group. Second, the primary endpoint was a biochemical surrogate over six weeks. Blood phenylalanine reduction is an accepted surrogate in this disorder, supported by the guideline's appraisal linking lifelong concentrations at or below 360  $\mu\text{mol/L}$  to better intellectual outcomes (Smith *et al.*, 2025), but a six-week biochemical endpoint is a considerable distance from the neurocognitive outcomes that justify treatment. Third, the trial was industry-sponsored, as are essentially all pivotal trials in this disorder, which is a structural feature of rare disease therapeutics rather than a specific criticism but which warrants explicit acknowledgement.

The extension study addresses the endpoint that matters most for daily life. In an interim analysis of the open-label phase 3 extension, 169 participants had received sepiapterin with a median age of 14.0 years and median exposure of 72.9 weeks, and 102 underwent formal dietary phenylalanine tolerance assessment; mean dietary phenylalanine intake increased from 27.6 mg/kg/day at baseline to 62.5 mg/kg/day at week 26, a least-squares mean increase of 36.4 mg/kg/day (van Spronsen *et al.*, 2026). A more than

twofold increase in tolerated dietary phenylalanine is a substantial and patient-relevant change. It is, however, an uncontrolled open-label observation, and dietary intake is a behaviourally mediated outcome susceptible to expectation effects in both participants and supervising clinicians. Corresponding dietetic guidance has been developed to support implementation (MacDonald *et al.*, 2026), which is appropriate, but the guidance necessarily rests on the same uncontrolled evidence.

The most informative recent addition is the first head-to-head comparison. AMPLIPHY was an international, two-part, open-label phase 3 study in participants aged two years and above, in which sepiapterin responders were randomised to crossover sequences of sepiapterin at 60 mg/kg/day and sapropterin at 20 mg/kg/day, the maximum licensed dose, for four weeks each with a 14-day washout (Gizewska *et al.*, 2026). Of 82 participants enrolled, 67 (81.7%) and 62 (75.6%) achieved reductions of at least 20% and 30% respectively in part 1. In the primary analysis set of 58 participants, least-squares mean reduction in blood phenylalanine was  $-437.0 \mu\text{mol/L}$  with sepiapterin and  $-256.6 \mu\text{mol/L}$  with sapropterin, a least-squares mean difference of  $-180.4 \mu\text{mol/L}$  (95% CI  $-229.5$  to  $-131.4$ ), with no new safety signals. This is the comparison the field required, and it supports superiority of sepiapterin over the highest approved sapropterin dose for phenylalanine lowering. Its limitations should be stated with equal clarity: the study was open-label, which matters for a behaviourally influenced outcome even when the primary measure is biochemical; participants were selected as sepiapterin responders in part 1, which structurally advantages sepiapterin in the crossover; the treatment periods were four weeks; and no cognitive, functional or quality-of-life outcome was assessed. The conclusion supported is that sepiapterin lowers blood phenylalanine more than sapropterin in sepiapterin-responsive individuals over four weeks. Whether that difference produces better long-term outcomes is not addressed by the trial and cannot be inferred from it.

### 6.3 Pegvaliase: Efficacy Without Genotype Dependence, at a Cost

Pegvaliase is pegylated recombinant phenylalanine ammonia lyase from *Anabaena variabilis*, which degrades phenylalanine to trans-cinnamic acid and ammonia independently of PAH. It is therefore the only approved pharmacological agent whose efficacy does not depend on residual enzyme function, and correspondingly the only one available to individuals with complete loss of function.

The pivotal evidence is a randomised discontinuation design. In part 2 of PRISM-2, adults already responding to pegvaliase with at least 20% phenylalanine reduction were randomised 2:1 to continue treatment or switch to matching placebo for eight weeks (Harding *et al.*, 2018). Sixty-six participants continued pegvaliase and 14 were allocated to each placebo group. At week 8, least-squares mean change in blood phenylalanine was  $949.8 \mu\text{mol/L}$  in the 20 mg/day placebo group and  $664.8 \mu\text{mol/L}$  in the 40 mg/day placebo group, compared with  $26.5 \mu\text{mol/L}$  in the pooled pegvaliase group. The magnitude of rebound on withdrawal is striking and constitutes strong evidence of pharmacological effect. The design nonetheless estimates the effect of withdrawal in established responders rather than the effect of initiation in an unselected population, and the two quantities are not interchangeable.

The wider phase 3 programme, reported alongside the discontinuation trial, documented sustained phenylalanine reduction across longer exposure together with the immunological profile that has since shaped clinical protocols (Thomas *et al.*, 2018). Final aggregated results from the PRISM programme give the more clinically useful picture. Across 261 adults with a mean of 36.6 months of treatment, 71.3%, 65.1% and 59.4% achieved blood phenylalanine of 600, 360 and  $120 \mu\text{mol/L}$  or below respectively, with substantial individual variation in dose requirement and in time to first response (Harding *et al.*, 2024). Achieving concentrations below  $120 \mu\text{mol/L}$  in nearly three-fifths of a treated adult population, many of whom had never achieved target control on diet, is the largest biochemical effect demonstrated by any approved therapy in this disorder. Against this stands the immunological burden: pegvaliase provokes predictable hypersensitivity reactions requiring induction, titration and maintenance protocols and clinical management pathways that were developed specifically for this purpose (Longo *et al.*, 2019). Comparative effectiveness analysis against medical nutrition therapy with and without sapropterin favours pegvaliase for phenylalanine control over the longer term (Burton *et al.*, 2024), but such analyses are observational and subject to confounding by indication, since patients offered pegvaliase are by definition those in whom prior management failed. Extension into adolescence is under active investigation, with multi-site experience now reported (Hollander *et al.*, 2026), and consensus recommendations for the management of early-treated adolescents and young adults address the wider question of how these agents are integrated into care across the transition period (Burton *et al.*, 2022).

### 6.4 Substrate Degradation in the Gut and Chaperones in Development

A conceptually distinct approach delivers phenylalanine-degrading activity to the gastrointestinal lumen rather than the circulation. Engineered *Escherichia coli* Nissle 1917 strains expressing phenylalanine ammonia lyase were evaluated in a 14-day open-label dose-escalation study in adults with screening phenylalanine above  $600 \mu\text{mol/L}$ , using the area under the curve of labelled phenylalanine after an isotopic challenge as the primary endpoint (Vockley *et al.*, 2023). Of 20 enrolled participants, 15 completed treatment. The second-generation strain reduced labelled phenylalanine area under the curve by 43% from baseline and fasting plasma phenylalanine by 40% (95% CI  $-52$  to  $-24$ ), with no serious adverse events or infections, although four participants discontinued because of adverse events. The pharmacological signal is genuine and the safety profile in this short exposure was acceptable. The evidence is nonetheless open-label, uncontrolled, of 14 days' duration and in 15 completers, and the primary endpoint was an isotopic pharmacodynamic measure rather than a clinical outcome. Supporting quantitative work modelling strain activity in healthy volunteers and patients strengthens the mechanistic account (Charbonneau *et al.*, 2021) without addressing the clinical question.

Chemical chaperones targeting PAH stability remain an unrealised opportunity. Compounds identified through the original pharmacological chaperone screen corrected the phenotype in mouse models more than a decade ago (Santos-Sierra *et al.*, 2012), building on the principle established in the initial identification of chaperone candidates (Pey *et al.*, 2008). No agent from this line

has reached approval. The most plausible explanation is that variant-specific stabilisation requires either a compound with broad activity across structurally diverse variants, which is pharmacologically demanding, or a portfolio of variant-specific compounds, which is commercially unattractive for a disorder in which most genotypes are rare. This is the same scalability problem that recurs in genome editing, encountered first in small-molecule form.

Table 3 compares the approved and investigational pharmacological options on the dimensions that determine clinical choice, with explicit attention to the design of the evidence rather than to headline effect size alone. Read across rows, the table shows that the field has produced increasingly sophisticated biochemical evidence without correspondingly advancing the measurement of clinical outcome; every pivotal trial listed used blood phenylalanine as its primary endpoint over a horizon of weeks, in a disorder whose morbidity accrues over decades.

**Table 3. Comparative appraisal of approved and investigational pharmacological therapies**

| Therapy                                                         | Mechanism                                                                               | Eligible population                                                                                | Strongest evidence and design                                                                 | Primary endpoint and horizon                                                    | Principal evidential limitation                                                                              | Supporting references                                           |
|-----------------------------------------------------------------|-----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Sapropterin                                                     | BH4 cofactor supplementation and pharmacological chaperone stabilisation of mutant PAH  | Cofactor-responsive individuals, largely those retaining folding-competent enzyme                  | Multinational observational registry, 576 participants, 69 sites, follow-up over a decade     | Dietary Phe tolerance and safety; years                                         | No untreated comparator; enrolment enriched for responders; response definitions vary between protocols      | Feillet <i>et al.</i> (2024, 2025); Muntau <i>et al.</i> (2019) |
| Sepiapterin                                                     | BH4 precursor with intracellular conversion, yielding broader response than sapropterin | Sepiapterin-responsive individuals, including some classic phenotypes and prior BH4 non-responders | Randomised, double-blind, placebo-controlled phase 3 (APHENITY), 34 sites, 13 countries       | Change in blood Phe; 6 weeks                                                    | Randomisation after responder enrichment; surrogate endpoint; short horizon; industry-sponsored              | Muntau <i>et al.</i> (2024); Lamb (2026)                        |
| Sepiapterin versus sapropterin                                  | Direct comparison of cofactor-based agents                                              | Sepiapterin responders aged $\geq 2$ years                                                         | Randomised open-label crossover phase 3 (AMPLIPHY), 62 randomised                             | Change in blood Phe; 4 weeks per period                                         | Open-label; selection for sepiapterin response precedes crossover; no functional or cognitive outcome        | Gizewska <i>et al.</i> (2026)                                   |
| Pegvaliase                                                      | Enzyme substitution with pegylated phenylalanine ammonia lyase, independent of PAH      | Adults with uncontrolled blood Phe despite prior management, irrespective of genotype              | Randomised discontinuation phase 3 (PRISM-2) plus final aggregated programme data, 261 adults | Change in blood Phe; 8 weeks (discontinuation) and 36.6 months mean (aggregate) | Estimates withdrawal effect in established responders; hypersensitivity burden requires structured induction | Harding <i>et al.</i> (2018, 2024); Longo <i>et al.</i> (2019)  |
| Engineered gut bacterium expressing phenylalanine ammonia lyase | Luminal degradation of dietary and enterorecirculated Phe                               | Adults with elevated blood Phe; investigational                                                    | Open-label dose-escalation phase 2, 20 enrolled and 15 completing                             | Labelled Phe area under the curve and fasting plasma Phe; 14 days               | Uncontrolled; very short exposure; small completer sample; pharmacodynamic endpoint                          | Vockley <i>et al.</i> (2023); Charbonneau <i>et al.</i> (2021)  |
| Chemical pharmacological chaperones                             | Variant-selective stabilisation of misfolded PAH                                        | Individuals with folding-defective variants; preclinical                                           | Mouse model correction following structure-guided identification                              | Blood Phe and enzyme stability; preclinical                                     | No clinical development to approval; variant specificity limits commercial and clinical scalability          | Pey <i>et al.</i> (2008); Santos-Sierra <i>et al.</i> (2012)    |

*Note.* BH4 = tetrahydrobiopterin; PAH = phenylalanine hydroxylase; Phe = phenylalanine; CI = confidence interval. Rows are not ranked; populations, comparators and endpoints differ sufficiently that direct comparison of effect magnitudes between rows is not warranted

## 7. Gene Addition, Transcript Replacement and Corrective Editing

### 7.1 Why Liver-Directed Correction Is Attractive and Why It Has Been Difficult

Phenylalanine hydroxylase deficiency has long been regarded as a favourable target for liver-directed intervention. The enzyme is expressed predominantly in hepatocytes, the pathology is metabolic rather than structural, the therapeutic biomarker is measurable

in a dried blood spot, and preclinical models are well established. Comprehensive reviews of the field describe a progression from viral gene addition through transcript replacement to corrective editing, with delivery technology rather than target biology as the principal rate-limiting factor (Martinez *et al.*, 2024; Yu & Jeong, 2025).

Gene addition using recombinant adeno-associated viral vectors faces a structural difficulty specific to this disorder. Because episomal vector genomes do not replicate, hepatocyte turnover progressively dilutes transgene expression, and this dilution is most rapid in the growing liver of a child, which is precisely the population in whom early intervention would yield the greatest benefit. The result is an uncomfortable trade-off in which the therapy is most durable in the patients who have already sustained the exposure it is intended to prevent. Clinical development in this disorder has proceeded, but peer-reviewed publication of clinical outcomes from liver-directed gene addition programmes has not kept pace with programme activity, and the available reviews describe these programmes without being able to appraise published patient-level results (Martinez *et al.*, 2024; Yu & Jeong, 2025). The absence of peer-reviewed clinical outcome data for an intervention in clinical development is itself a finding, and one that argues for caution in claims about the readiness of gene addition for this indication.

## 7.2 Transcript Replacement: Effective in Principle, Chronic in Practice

Delivery of messenger RNA encoding human PAH in lipid nanoparticles circumvents the durability problem by abandoning permanence altogether. Intravenous administration of a lipid nanoparticle-encapsulated full-length human *PAH* transcript generated high levels of the protein in hepatocytes and restored phenylalanine metabolism in the *Pah<sup>enu2</sup>* mouse, establishing proof of principle (Perez-Garcia *et al.*, 2022). A parallel approach delivered phenylalanine ammonia lyase transcript rather than *PAH*, achieving hepatic catabolism and clearance of phenylalanine in the same model (Diaz-Trelles *et al.*, 2022). Systematic preclinical characterisation of an investigational transcript therapy alongside comparable products for propionic and methylmalonic acidemia demonstrated pharmacokinetic and pharmacodynamic responses in the respective murine models and supported allometric scaling to predicted first-in-human doses (Baek *et al.*, 2024).

The strengths of this evidence are its mechanistic clarity and the transparency of the dose-prediction methodology. The limitations are intrinsic to the modality. Transcript replacement requires repeated intravenous administration indefinitely, which converts a genetic disorder into a chronic infusion-dependent condition, raises questions about lifetime immunogenicity of both the lipid carrier and the expressed protein, and carries a cumulative cost burden that no current health system has been asked to sustain for this indication. Transcript replacement is best understood as a bridging modality rather than as a competitor to permanent correction, and its most plausible clinical niche may be neonatal or peri-diagnostic use pending definitive intervention.

## 7.3 Adenine Base Editing: The Strongest Preclinical Evidence in the Field

Adenine base editing converts adenine–thymine to guanine–cytosine base pairs without generating a double-strand break, and because the most frequent pathogenic *PAH* variants are transition substitutions amenable to reversal by this chemistry, the disorder is unusually well suited to the approach.

The evidence has accumulated with notable internal consistency. Correction of the c.842C>T (p.Pro281Leu) variant was demonstrated in a prime-edited hepatocyte line and in a variant-humanised mouse model, with delivery of ABE8.8 messenger RNA and either of two guide RNAs in lipid nanoparticles producing complete and durable normalisation of blood phenylalanine within 48 hours of treatment (Brooks, Carrasco, *et al.*, 2023). An analogous strategy directed at c.1222C>T (p.Arg408Trp), the most frequent pathogenic variant worldwide, achieved sufficient hepatic editing to fully normalise blood phenylalanine within the same interval, with the authors explicitly noting potential limitations relative to other editing approaches (Brooks *et al.*, 2024). Independent work using dual adeno-associated viral vectors to deliver a near-protospacer-adjacent-motif-less editor variant to the *Pah<sup>R408W</sup>* mouse achieved 19.1% to 34.6% A-to-G editing at the pathogenic site with minimal bystander editing following a single injection, reduced blood phenylalanine below the 360 µmol/L threshold, restored weight and coat colour, and normalised or near-normalised brain monoamine neurotransmitters and metabolites, with the strongest effects when treatment was given neonatally (Yin *et al.*, 2025).

Two further studies extend the evidence in directions that matter clinically. Screening of editor and guide combinations against the next four most recurrent variants established efficient corrective editing in vitro for three candidates, characterised the off-target profile of each combination, and demonstrated efficient hepatic correction in humanised mouse models for two variants using adeno-associated viral delivery (Quigley *et al.*, 2025). Separately, treatment of humanised P281L mice at four weeks of age with lipid nanoparticle-delivered editor produced significant reductions in phenylalanine in both plasma and cortex, increases in cerebral amino acid and neurotransmitter concentrations, and improvement in abnormal motor phenotypes assessed behaviourally into adulthood (Gautier *et al.*, 2026).

Taken together, this body of work supports a strong conclusion: in variant-humanised models carrying the actual human pathogenic allele, a single administration of an adenine base editor produces rapid, complete and durable normalisation of the biochemical phenotype, with demonstrable cerebral and behavioural consequences. That is a materially stronger evidential position than any previous therapeutic modality reached before first-in-human study. Four qualifications nonetheless apply. Murine hepatocyte biology, liver growth kinetics and immune responses differ from human, and editing efficiencies do not transfer directly. Behavioural improvement in mice is not cognitive benefit in humans. Durability demonstrated over months in a short-lived species does not establish durability over decades. And no peer-reviewed clinical outcome from corrective editing in this disorder has been published.

## 7.4 Prime Editing: Greater Flexibility, Lower Efficiency

Prime editing writes new sequence from a template carried on an extended guide RNA and is not restricted to transition substitutions, making it applicable to variant classes that base editing cannot address. Early demonstration in the liver established *in vivo* feasibility in a metabolic disease model (Böck *et al.*, 2022). Applied to c.1222C>T, adeno-associated viral delivery of prime editing components reproducibly normalised blood phenylalanine in humanised mice, with up to 52% whole-liver corrective editing, and the same study documented through real-world data that most individuals carrying at least one c.1222C>T allele experience chronic severe phenylalanine elevation despite dietary and medical intervention (Brooks, Whittaker, *et al.*, 2023). That combination of clinical need and preclinical efficacy is the clearest articulation available of why corrective editing is being pursued.

Subsequent work addressed the exposure problem created by persistent editor expression. A dual-delivery strategy in which the prime editing guide RNA is expressed continuously from an adeno-associated viral vector while the prime editor is delivered transiently as nucleoside-modified messenger RNA in lipid nanoparticles achieved 26.2% editing with PEmax and 47.4% with PE7 at a reporter locus after a single 2 mg/kg dose; targeting the *Pah*<sup>enu2</sup> variant, correction reached 4.3% with PEmax and 20.7% with PE7 after three doses, reducing blood phenylalanine from above 1,500 µmol/L to below the 360 µmol/L threshold (Rothgangl *et al.*, 2025). The comparison between the reporter locus and the disease locus is instructive: efficiency at a therapeutic target was substantially lower than at an optimised reporter, and required repeated dosing. Prime editing therefore trades breadth of applicability against efficiency, and the trade-off is currently unfavourable relative to base editing for the transition variants that both can address.

## 7.5 Delivery, Off-Target Editing and the Limits of Current Safety Characterisation

The safety question for corrective editing has three components that are often conflated: off-target editing at unintended genomic sites, bystander editing of adjacent bases within the target window, and consequences of the delivery vehicle itself.

Off-target characterisation in this disorder has become substantially more rigorous. Comprehensive and variant-aware specificity profiling of clinical lead guide RNAs for the P281L and R408W variants, together with systematic screening of hybrid guide RNAs incorporating DNA nucleotide substitutions, demonstrated markedly reduced off-target and bystander editing in human hepatocytes, and selected hybrid guides formulated with editor messenger RNA in lipid nanoparticles reverted disease phenotypes in humanised mice while simultaneously increasing on-target and reducing off-target and bystander editing (Whittaker *et al.*, 2025). The significance of this work lies less in the specific guides than in the demonstration that specificity is engineerable rather than fixed, and in the use of variant-aware profiling, which accounts for the fact that an individual's own genetic variation alters their off-target landscape. The broader methodological literature nonetheless documents that off-target assessment methods remain incompletely standardised and that translation of *in vitro* specificity profiles to clinical risk estimates is unresolved (Kalter *et al.*, 2025).

Dose-response relationships within the delivery system introduce a further constraint. In a reporter mouse model, increasing guide RNA dose to 4 mg/kg enhanced editing up to 3.3-fold across tissues relative to a standard 1 mg/kg dose, achieving 63% on-target editing in liver, but bystander editing increased significantly at higher guide doses, particularly in the most efficiently edited tissues (Birkenshaw *et al.*, 2026). On-target efficiency and bystander fidelity are therefore coupled through dose, and the therapeutic window must be defined empirically for each editor and guide combination rather than assumed. Commentary within the field has additionally emphasised that the number of doses required, the choice between viral and non-viral delivery, and the resulting exposure profile are as consequential for safety as the editor chemistry itself (Nicosia, 2026).

The most serious unresolved problem is not technical but structural. With more than a thousand recorded pathogenic *PAH* variants, an allele-specific correction strategy addresses only those individuals carrying the targeted variant. The most frequent variant accounts for approximately 22% of alleles and the three most frequent for roughly a third (Hillert *et al.*, 2020), so even successful development of editors for the five most common variants would leave a substantial minority of patients, disproportionately those with rare genotypes, without access. This scalability problem has been stated explicitly in the recent literature and has no current solution within the allele-specific paradigm (Nicosia, 2026). Approaches that restore function without correcting the specific variant, including transgene insertion at a safe harbour or promoterless targeting to an endogenous locus, would sidestep the problem but reintroduce the durability and integration-site questions that corrective editing was intended to avoid.

Table 4 summarises the principal preclinical genome-editing and transcript-based studies in this disorder, presenting the model, delivery route, efficiency achieved and the specific limitation of each. The table is restricted to studies in which the disease-relevant locus was targeted in a living animal, since *in vitro* editing efficiencies have historically overestimated *in vivo* performance. Two patterns are visible across the rows: efficiency at disease loci is consistently lower than at optimised reporter loci, and every study that measured cerebral or behavioural outcome did so in mice, which is where the translational gap is widest.

## 8. Scalability, Measurement and Access as Constraints on Translation

Three constraints cut across the modalities discussed and will determine which of them reach patients.

The first is scalability under genetic heterogeneity, discussed above in its editing-specific form but general in its reach. Cofactor-based agents depend on folding-competent residual enzyme, chemical chaperones on variant-specific stabilisation, and corrective editing on allele identity. Only enzyme substitution and substrate degradation are genotype-agnostic, and both impose either immunological burden or chronic administration. The disorder's allelic architecture, in which most genotypes are compound heterozygous and most distinct genotypes are observed once (Hillert *et al.*, 2020), places a structural ceiling on any precision strategy. Recognising this does not argue against precision approaches; it argues for parallel development of genotype-agnostic options and against presenting allele-specific correction as a solution for the disease as a whole.

**Table 4. Preclinical evidence for transcript replacement and corrective genome editing in phenylalanine hydroxylase deficiency**

| <b>Approach and target</b>                                                   | <b>Model</b>                              | <b>Delivery</b>                                                                 | <b>Principal efficacy result</b>                                                                                               | <b>Principal limitation</b>                                                            | <b>Supporting references</b>            |
|------------------------------------------------------------------------------|-------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------------------------------|
| Transcript replacement, human <i>PAH</i> mRNA                                | <i>Pah<sup>enu2</sup></i> mouse           | Intravenous lipid nanoparticle                                                  | High hepatocyte expression of human PAH with restoration of Phe metabolism                                                     | Requires indefinite repeat dosing; no permanence                                       | Perez-Garcia <i>et al.</i> (2022)       |
| Transcript replacement, phenylalanine ammonia lyase mRNA                     | Phenylketonuria mouse                     | Lipid nanoparticle to liver                                                     | Hepatic catabolism and clearance of Phe                                                                                        | Non-native enzyme; repeat dosing; immunogenicity unassessed long term                  | Diaz-Trelles <i>et al.</i> (2022)       |
| Transcript replacement, investigational product with translational modelling | Murine disease models                     | Lipid nanoparticle                                                              | Pharmacokinetic and pharmacodynamic responses supporting allometrically scaled first-in-human doses                            | Dose prediction is modelled, not observed in humans                                    | Baek <i>et al.</i> (2024)               |
| Adenine base editing, c.842C>T (P281L)                                       | Variant-humanised mouse                   | ABE8.8 mRNA with guide RNA in lipid nanoparticle                                | Complete and durable normalisation of blood Phe within 48 h                                                                    | Single variant; murine hepatocyte biology differs from human                           | Brooks, Carrasco, <i>et al.</i> (2023)  |
| Adenine base editing, c.1222C>T (R408W)                                      | Variant-humanised mouse                   | Editor mRNA and guide RNA in lipid nanoparticle                                 | Full normalisation of blood Phe within 48 h                                                                                    | Authors note limitations relative to alternative editing approaches                    | Brooks <i>et al.</i> (2024)             |
| Adenine base editing, <i>Pah<sup>R408W</sup></i> with PAM-relaxed editor     | <i>Pah<sup>R408W</sup></i> mouse          | Dual adeno-associated virus, single injection                                   | 19.1%–34.6% on-target editing; blood Phe below 360 µmol/L; brain neurotransmitters normalised, strongest when neonatal         | Viral delivery raises integration and immunity questions; rodent behavioural endpoints | Yin <i>et al.</i> (2025)                |
| Adenine base editing, next four recurrent variants                           | Humanised mouse models                    | Adeno-associated virus                                                          | Efficient corrective editing <i>in vivo</i> for two of the variants assessed                                                   | Addresses only recurrent alleles; rare genotypes remain unaddressed                    | Quigley <i>et al.</i> (2025)            |
| Adenine base editing with cerebral and behavioural readout                   | Humanised P281L mouse, treated at 4 weeks | Lipid nanoparticle                                                              | Reduced plasma and cortical Phe; increased cerebral amino acids and neurotransmitters; improved motor phenotype into adulthood | Behavioural improvement in mice is not cognitive benefit in humans                     | Gautier <i>et al.</i> (2026)            |
| Prime editing, c.1222C>T                                                     | Humanised mouse                           | Adeno-associated virus                                                          | Complete normalisation of blood Phe; up to 52% whole-liver corrective editing                                                  | Lower efficiency than base editing at the same locus                                   | Brooks, Whittaker, <i>et al.</i> (2023) |
| Prime editing, <i>Pah<sup>enu2</sup></i> , transient editor exposure         | <i>Pah<sup>enu2</sup></i> mouse           | Guide RNA from adeno-associated virus with editor as mRNA in lipid nanoparticle | 20.7% correction with PE7 after three doses; blood Phe reduced from >1,500 to <360 µmol/L                                      | Repeat dosing required; efficiency at disease locus well below reporter locus          | Rothgangl <i>et al.</i> (2025)          |

| Approach and target                   | Model                                        | Delivery                                                 | Principal efficacy result                                                                                     | Principal limitation                                                     | Supporting references                                       |
|---------------------------------------|----------------------------------------------|----------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| Guide RNA engineering for specificity | Human hepatocytes and humanised mouse models | Lipid nanoparticle with editor mRNA                      | Reduced off-target and bystander editing with increased on-target editing; phenotype reversion <i>in vivo</i> | Specificity profiling methods not yet standardised across laboratories   | Whittaker <i>et al.</i> (2026); Kalter <i>et al.</i> (2025) |
| Guide RNA dose-response               | Reporter mouse                               | Lipid nanoparticle co-encapsulating editor and guide RNA | 63% on-target hepatic editing at 4 mg/kg guide dose, 3.3-fold above standard dose                             | Bystander editing rises with guide dose, coupling efficiency to fidelity | Birkenshaw <i>et al.</i> (2026)                             |

Note. PAM = protospacer adjacent motif; PAH = phenylalanine hydroxylase; Phe = phenylalanine. All efficacy results are from animal models; no peer-reviewed clinical outcome data from corrective editing in this disorder were identified within the search period

The second is measurement. Every pivotal trial in the modern era has used blood phenylalanine as its primary endpoint over weeks, while the outcomes that justify treatment accrue over decades and are cognitive, psychiatric, skeletal and social. The surrogate is defensible, and the guideline's evidence appraisal supports the link between lifelong concentrations at or below 360  $\mu\text{mol/L}$  and better intellectual outcomes (Smith *et al.*, 2025). It is nonetheless a surrogate, and the field has not developed the disease-specific, validated, longitudinally sensitive outcome instruments that would allow therapies to be compared on what matters to patients. The scoping review of quality of life identified exactly this deficiency (Remor *et al.*, 2024), and the meta-analytic literature demonstrates that cognitive outcomes have been measured with instruments varying so widely that pooling required extraction across 19 separate functions probed by an average of 12 different measures each (Romani *et al.*, 2022). Dietary phenylalanine tolerance, now used as a primary endpoint in extension studies (van Spronsen *et al.*, 2026), is more patient-relevant than blood concentration but is behaviourally mediated and has been measured only in open-label settings.

The third is access. Newborn screening, the precondition for all treatment, remains internationally variable. A survey of screening laboratories across 23 countries found broad agreement on sampling window, analytical platform and the 360  $\mu\text{mol/L}$  threshold for initiating dietary therapy, alongside striking differences in other aspects of practice (Trampuž *et al.*, 2025). Long-running national programmes continue to report experience from settings where coverage has historically been incomplete (Kumar & Prabha, 2026). Where screening is absent or delayed, no therapy discussed in this review is relevant, because irreversible neurological injury precedes diagnosis. The economic dimension compounds this. Work developing a face-validated conceptual model structure for economic evaluation in the disorder through expert elicitation indicates that the field is still constructing the analytical apparatus required to evaluate these therapies economically (Muntau *et al.*, 2026), which means that health technology assessment of high-cost therapies is proceeding without a settled modelling framework. The disparity between the cost of a single administration of a corrective editing therapy and the cost of newborn screening in a low-resource setting is the sharpest equity question the field faces, and it is largely absent from the mechanistic literature.

Table 5 sets out the principal research gaps identified in this synthesis, the reason current evidence is inadequate to close them, and the design features that would materially advance each. The gaps are ordered by the immediacy with which they constrain clinical decision-making rather than by scientific interest, and several could be addressed with existing methods and modest resources, which makes their persistence a matter of research priority rather than technical impossibility.

**Table 5. Principal research gaps, evidential inadequacy and design requirements**

| Research gap                                              | Why current evidence is inadequate                                                                                                                                   | Design or methodological requirement                                                                                                            | Priority  | Supporting references                                                                                                 |
|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------------------------------------------------------------------------------------------------------------------|
| Whether adult blood Phe targets improve long-term outcome | Meta-analytic data relate outcome more strongly to childhood than concurrent exposure; the only randomised challenge lasted four weeks and tested a preserved domain | Multi-year randomised or stepped-wedge comparison of target strategies in adults, with prespecified processing-speed and functional endpoints   | Immediate | Romani <i>et al.</i> (2025); Trepp <i>et al.</i> (2024); Burgard <i>et al.</i> (2026); Langeveld <i>et al.</i> (2026) |
| Disentangling morbidity of disease from morbidity of diet | Comorbidity and bone density literature cannot separate Phe exposure from chronic protein restriction                                                                | Cohorts stratified by therapy modality, comparing pharmacologically liberalised against dietarily restricted patients at equivalent Phe control | Immediate | Whitehall <i>et al.</i> (2024); Rocha <i>et al.</i> (2024)                                                            |

| Research gap                                                         | Why current evidence is inadequate                                                                                           | Design or methodological requirement                                                                                                          | Priority    | Supporting references                                                                             |
|----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|-------------|---------------------------------------------------------------------------------------------------|
| Valid patient-relevant outcome measures                              | Cognitive measures are non-comparable across studies; quality-of-life instruments are largely non-disease-specific           | Development and psychometric validation of disease-specific instruments with demonstrated longitudinal sensitivity                            | Immediate   | Remor <i>et al.</i> (2024); Romani <i>et al.</i> (2022)                                           |
| Whether cofactor superiority in Phe lowering yields clinical benefit | Head-to-head comparison used a four-week open-label crossover with a biochemical endpoint                                    | Extended blinded comparison with dietary tolerance, neurocognitive and quality-of-life endpoints                                              | Near term   | Gizewska <i>et al.</i> (2026); van Spronsen <i>et al.</i> (2026)                                  |
| Durability and long-term safety of corrective editing                | Durability demonstrated over months in mice; human hepatocyte turnover and immune responses differ                           | First-in-human studies with prespecified multi-year follow-up, serial off-target surveillance and registry linkage                            | Near term   | Brooks, Carrasco, <i>et al.</i> (2023); Gautier <i>et al.</i> (2026); Kalter <i>et al.</i> (2025) |
| Strategies for patients with rare genotypes                          | Allele-specific editing addresses recurrent variants only; most distinct genotypes occur once                                | Genotype-agnostic correction strategies, or platform regulatory pathways permitting efficient development of variant-specific editors         | Near term   | Hillert <i>et al.</i> (2020); Nicosia (2026); Quigley <i>et al.</i> (2025)                        |
| Optimal age at definitive intervention                               | Preclinical benefit is greatest with neonatal treatment, but risk tolerance is lowest in neonates                            | Prospective comparison of intervention timing with neurodevelopmental endpoints, preceded by explicit ethical frameworks for neonatal editing | Longer term | Yin <i>et al.</i> (2025); Gautier <i>et al.</i> (2026)                                            |
| Equitable global access                                              | Screening coverage and confirmatory practice vary internationally; economic modelling frameworks are still being constructed | Implementation research linking screening coverage to outcome, and economic evaluation using an agreed model structure                        | Longer term | Trampuž <i>et al.</i> (2025); Kumar & Prabha (2026); Muntau <i>et al.</i> (2026)                  |

*Note. Phe = phenylalanine. Priority reflects the immediacy with which the gap constrains clinical decision-making, not scientific importance*

## 9. Future Directions

The priorities below are ordered by how directly they constrain decisions that clinicians and patients must make now, and each is specified in terms of the design that would resolve it rather than as a general call for more research.

**Resolving the adult treatment target.** The most consequential open question in routine practice is whether adults with well-controlled childhood exposure derive sufficient benefit from maintaining blood phenylalanine at or below 360  $\mu\text{mol/L}$  to justify the burden of doing so. The current evidence cannot settle it. Guideline appraisal supports the target on the basis of lifelong exposure data (Smith *et al.*, 2025), experienced clinicians have publicly disputed aspects of that appraisal (Burgard *et al.*, 2026), and a multinational group has argued for individualised goal-setting through shared decision making (Langeveld *et al.*, 2026). The only randomised challenge in adults ran for four weeks and assessed a domain that meta-analytic evidence identifies as relatively preserved (Romani *et al.*, 2025; Trepp *et al.*, 2024). What is required is a multi-year randomised or stepped-wedge comparison of target strategies in adults with documented good childhood control, powered for processing-speed measures rather than accuracy measures, with prespecified secondary endpoints covering psychiatric symptoms, bone density, employment and treatment burden, and with adherence measured rather than assumed. Such a trial is feasible within existing metabolic networks and would not require new technology. Its absence is a failure of research prioritisation rather than of research capacity.

**Separating the disease from its treatment.** The somatic comorbidity literature cannot distinguish the consequences of phenylalanine exposure from those of chronic restriction of natural protein, and the bone density evidence is the clearest example (Rocha *et al.*, 2024; Whitehall *et al.*, 2024). The arrival of pharmacological agents that permit substantial dietary liberalisation creates, for the first time, a natural comparison: patients achieving equivalent phenylalanine control through pharmacological

liberalisation versus through dietary restriction. Prospective cohorts stratified by modality and matched on achieved phenylalanine concentration, with bone densitometry, nutritional biomarkers, renal function and inflammatory markers assessed over at least five years, would apportion the comorbidity that current designs conflate. The dietary tolerance data emerging from cofactor extension studies indicate that sufficient numbers of liberalised patients now exist for such cohorts to be assembled (van Spronsen *et al.*, 2026).

**Building outcome measures worth trialling against.** No therapy can be shown to improve what is not validly measured, and the field currently lacks disease-specific, longitudinally sensitive instruments for the outcomes that motivate treatment. The scoping review of quality of life documented the reliance on generic instruments (Remor *et al.*, 2024), and the meta-analytic literature required extraction across nineteen functions probed by an average of twelve measures each, which is a direct index of non-comparability (Romani *et al.*, 2022). A coordinated programme to develop and psychometrically validate a core outcome set, covering processing speed, executive function, psychiatric symptoms, treatment burden and social participation, with demonstrated sensitivity to change over one to two years, should precede rather than follow the next generation of trials. Without it, editing trials will report blood phenylalanine and the field will be no better placed to judge them than it is to judge cofactor agents now.

**Extending the cofactor comparison beyond biochemistry.** The head-to-head comparison of sepiapterin and sapropterin established biochemical superiority over four weeks in an open-label crossover among sepiapterin responders (Gizewska *et al.*, 2026). The clinically decisive question is whether that biochemical margin produces greater dietary tolerance, better neurocognitive function or improved quality of life over a period long enough for such differences to emerge. An extended comparison, blinded where formulation permits, with dietary phenylalanine tolerance as a co-primary endpoint alongside a validated functional measure, and with at least twelve months of exposure, would answer it. Responder-enrichment designs should be reported with the unenriched response rate given equal prominence, since that is the figure clinicians need when deciding whom to trial on therapy.

**Defining the evidentiary threshold for first-in-human corrective editing.** Preclinical base editing evidence in this disorder is now strong and internally consistent across independent groups (Brooks, Carrasco, *et al.*, 2023; Gautier *et al.*, 2026; Quigley *et al.*, 2025; Yin *et al.*, 2025). What has not been articulated is what a first-in-human programme must demonstrate, and over what period, to justify replacing an effective if burdensome chronic therapy with a permanent genomic alteration. Prespecification should cover the duration of follow-up, which should be measured in decades rather than years given that the intervention is irreversible; the method and frequency of off-target surveillance, using variant-aware profiling that accounts for participants' own genetic backgrounds (Whittaker *et al.*, 2026); the handling of bystander editing, whose relationship to on-target efficiency is dose-dependent and must therefore be characterised across the intended dose range rather than at a single dose (Birkenshaw *et al.*, 2026); and the registry infrastructure required to detect late effects across small, internationally dispersed cohorts. Standardisation of off-target assessment across laboratories remains unresolved and is a prerequisite rather than a refinement (Kalter *et al.*, 2025).

**Addressing the rare-genotype majority.** Allele-specific correction leaves untreated those patients whose genotypes occur too infrequently to support dedicated development, and the allelic architecture of the disorder ensures that this group is substantial (Hillert *et al.*, 2020; Nicosia, 2026). Two routes merit parallel investment. The first is genotype-agnostic functional restoration, including targeted transgene integration at a defined locus, which would require resolution of the durability and insertion-site questions that corrective editing avoids. The second is regulatory and infrastructural: platform designation pathways in which the editor chemistry and delivery system are assessed once and individual guide RNAs are evaluated through an abbreviated process would make variant-specific development tractable for rare alleles. Progress on the second may ultimately matter more than progress on the first, and it requires engagement between investigators, regulators and patient organisations rather than further laboratory work.

**Determining the optimal age of definitive intervention.** Preclinical evidence consistently indicates greater benefit from earlier treatment, with neonatal administration producing the most complete normalisation of cerebral neurotransmitter profiles (Yin *et al.*, 2025) and treatment at four weeks in mice nonetheless improving cerebral biochemistry and motor phenotype into adulthood (Gautier *et al.*, 2026). Clinical risk tolerance runs in the opposite direction, since irreversible intervention in a neonate who would do well on established therapy is difficult to justify. Resolving this tension requires explicit ethical frameworks developed in advance of trials, prospective comparison of intervention timing with neurodevelopmental endpoints, and honest communication that the population in whom benefit is greatest is also the population in whom uncertainty is greatest.

**Connecting therapeutic advance to screening reality.** None of the above is relevant to a child whose disorder is not detected. International variation in screening practice persists after six decades of programme operation (Trampuž *et al.*, 2025), and national programmes continue to report on the consolidation of coverage (Kumar & Prabha, 2026). Implementation research linking screening coverage and time to treatment initiation with neurodevelopmental outcome, conducted in settings where coverage is incomplete, would generate evidence of greater aggregate benefit than any refinement of therapy in settings where screening is universal. Economic evaluation of high-cost therapies should proceed using an agreed model structure of the kind currently being constructed (Muntau *et al.*, 2026), and should incorporate the opportunity cost of those therapies relative to screening expansion.

## 10. Conclusions

This review set out to examine whether the conceptual framework inherited from the first era of phenylketonuria treatment remains adequate for the therapeutic landscape now emerging, and to assess what the available evidence supports across dietary, pharmacological and genomic approaches.

The classical phenotypic categories retain descriptive value at population level but perform inconsistently for individual patients, and their thresholds are conventional rather than biological. Genotype adds real predictive information, particularly for common alleles and for anticipating cofactor responsiveness, but the allelic architecture of the disorder, in which most genotypes are compound heterozygous and a majority of distinct genotypes have been observed only once, places a ceiling on genotype-based prediction that no current model overcomes. The defensible position is that phenotype, genotype, functional annotation and measured cofactor response each contribute partial information, that none is sufficient alone, and that therapeutic selection still requires biochemical confirmation. Claims that any single framework now determines treatment overstate what has been demonstrated.

Residual morbidity in early and continuously treated disease is real, but its magnitude and attribution are more uncertain than advocacy on either side acknowledges. The cognitive effect is approximately half a standard deviation on average, structured rather than global, concentrated in speed rather than accuracy, and related more securely to childhood than to concurrent adult exposure. Somatic comorbidity is documented but confounded between the disorder and the diet used to treat it. Quality-of-life effects are reported through instruments not designed for this population. Each of these limitations weakens the chain of inference from residual burden to the justification for higher-risk intervention, and that chain is the principal argument advanced for advancing to definitive therapy.

The pharmacological landscape has genuinely improved. A cofactor precursor with a broader response population than its predecessor has been approved on the basis of a placebo-controlled trial and has now been compared directly with that predecessor, with superior phenylalanine lowering demonstrated over four weeks in responders. Enzyme substitution achieves the largest biochemical effect of any approved therapy and is the only genotype-independent option, at the cost of a substantial immunological burden. The common weakness across this literature is that every pivotal trial used a biochemical surrogate over a horizon of weeks in a disorder whose morbidity accrues over decades, and that the field has not built the outcome measures that would permit these agents to be compared on what matters to patients.

Corrective genome editing stands on the strongest preclinical foundation any modality in this disorder has achieved. Independent groups have demonstrated that a single administration of an adenine base editor produces rapid, complete and durable normalisation of blood phenylalanine in mice carrying human pathogenic alleles, with accompanying improvement in cerebral amino acid and neurotransmitter concentrations and in motor phenotype. That is a stronger position than gene addition or transcript replacement reached before clinical study. It is not, however, clinical evidence, and no peer-reviewed clinical outcome from corrective editing in this disorder was identified. Between the preclinical achievement and clinical adoption lie three problems that are not primarily technical: the absence of outcome measures against which benefit could be demonstrated, the absence of a prespecified evidentiary threshold for replacing effective chronic therapy with irreversible genomic alteration, and the structural exclusion of patients whose genotypes are too rare to support allele-specific development.

The broader significance of this synthesis is that the disorder's therapeutic problem has changed character. For six decades the question was how to lower phenylalanine. Multiple means of doing so now exist, with more in development, and the binding constraint has shifted to knowing which outcomes justify which risks, in whom, and at what age. That question is answered by measurement and by trial design, not by further mechanistic advance, and the field's present investment is weighted heavily towards the latter.

## 11. Limitations

This review is a narrative synthesis and carries the limitations intrinsic to that design. Study selection was purposive rather than exhaustive, and although the search was structured and documented, and appraisal criteria were applied explicitly, the possibility of selection bias in which studies were retrieved and interpretive bias in how they were weighed cannot be excluded. A different reviewer applying the same criteria to the same literature could reasonably assemble a somewhat different evidence base and reach different emphases, particularly on the contested question of adult treatment targets, where the published positions are themselves divided.

Access to subscription-indexed bibliographic databases was not available, and searching was confined to openly accessible scholarly sources and registries. Locally published or regionally indexed studies that are poorly represented in those sources may therefore be under-represented, and this is most likely to affect evidence from settings where screening coverage and treatment access differ most from the high-income contexts that dominate the literature. Eligibility was restricted to publications in English, which compounds the same risk.

The evidence base is geographically uneven in a manner that constrains generalisation. Trials, registries and cohort studies are concentrated in Europe, North America and a small number of other high-income settings, while the disorder's highest prevalence estimates and its most severe phenotype distributions occur partly in regions contributing little primary evidence. Statements about treatment effect should be read as applying to the populations studied.

Heterogeneity limited what could be synthesised. Cognitive outcomes have been measured with instruments that are not comparable across studies, dietary interventions differ in composition, duration and supervision, response definitions for cofactor therapies vary between protocols, and preclinical editing studies differ in model, delivery route and assay for off-target effects. Quantitative synthesis was therefore neither attempted nor appropriate, and no pooled effect estimates are reported. Comparisons between therapies in this review are qualitative and directional; they should not be read as establishing rank order.

Several specific weaknesses in the underlying evidence propagate into this review. Pivotal trials have been short relative to the chronicity of the disorder and have used biochemical surrogates. Responder-enrichment designs are common, so randomised estimates frequently describe effects in selected populations rather than in patients presenting for treatment. Observational and registry evidence is subject to confounding by indication that no analytical adjustment fully removes. Almost all pivotal trials have been industry-sponsored, which is unavoidable in rare disease but affects the balance of what is studied and published. Publication bias is likely in the preclinical editing literature in particular, where negative results are rarely reported and where the same small number of laboratories contributes much of the evidence.

Finally, this is a rapidly moving field, and the account given here is bounded by its search period. Corrective editing programmes are in active development, and clinical results that would materially alter the balance of this assessment may appear shortly after the period covered. The conclusions drawn about the readiness of these approaches should be understood as provisional to a degree that does not apply to the sections addressing dietary and pharmacological management.

## Consent

It is not applicable.

## Ethical Approval

It is not applicable.

## Declaration of AI Use

This manuscript was prepared through the intellectual contributions of the author(s) to the study design, data analysis, interpretation of results, and scholarly content. Grammarly and ChatGPT were used solely to assist with grammatical refinement and reference formatting during manuscript revision. The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

## Competing Interests

Authors have declared that no competing interests exist.

## References

- Baek, R., Coughlan, K., Jiang, L., Liang, M., Ci, L., Singh, H., Zhang, H., Kaushal, N., Rajlic, I. L., Van, L., Dimen, R., Cavedon, A., Yin, L., Rice, L., Frassetto, A., Guey, L., Finn, P., & Martini, P. G. V. (2024). Characterizing the mechanism of action for mRNA therapeutics for the treatment of propionic acidemia, methylmalonic acidemia, and phenylketonuria. *Nature Communications*, 15(1), 3804. <https://doi.org/10.1038/s41467-024-47460-9>
- Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4(1), 5. <https://doi.org/10.1186/s41073-019-0064-8>
- Birkenshaw, A., Thomson, T., Truong, M. P., Komaki, Y., Ramsden, N., Timpano, A., Huang, C., Blakney, A. K., Kurek, D. Z., Kulkarni, J., & Ross, C. J. D. (2026). sgRNA amount is a limiting factor in adenine base editing using RNA LNPs. *Molecular Therapy Nucleic Acids*, 37(2), 102909. <https://doi.org/10.1016/j.omtn.2026.102909>
- Blau, N. (2016). Genetics of phenylketonuria: Then and now. *Human Mutation*, 37(6), 508–515. <https://doi.org/10.1002/humu.22980>
- Böck, D., Rothgangl, T., Villiger, L., Schmidheini, L., Matsushita, M., Mathis, N., Ioannidi, E., Rimann, N., Grisch-Chan, H. M., Kreutzer, S., Kontarakis, Z., Kopf, M., Thöny, B., & Schwank, G. (2022). *In vivo* prime editing of a metabolic liver disease in mice. *Science Translational Medicine*, 14(636), eabl9238. <https://doi.org/10.1126/scitranslmed.abl9238>
- Bratkovic, D., Margvelashvili, L., Tchan, M. C., Nisbet, J., & Smith, N. (2022). PTC923 (sepiapterin) lowers elevated blood phenylalanine in subjects with phenylketonuria: A phase 2 randomized, multi-center, three-period crossover, open-label, active controlled, all-comers study. *Metabolism*, 128, 155116. <https://doi.org/10.1016/j.metabol.2021.155116>
- Brooks, D. L., Carrasco, M. J., Qu, P., Peranteau, W. H., Ahrens-Nicklas, R. C., Musunuru, K., Alameh, M.-G., & Wang, X. (2023). Rapid and definitive treatment of phenylketonuria in variant-humanized mice with corrective editing. *Nature Communications*, 14(1), 3451. <https://doi.org/10.1038/s41467-023-39246-2>
- Brooks, D. L., Whittaker, M. N., Qu, P., Musunuru, K., Ahrens-Nicklas, R. C., & Wang, X. (2023). Efficient *in vivo* prime editing corrects the most frequent phenylketonuria variant, associated with high unmet medical need. *The American Journal of Human Genetics*, 110(12), 2003–2014. <https://doi.org/10.1016/j.ajhg.2023.10.005>
- Brooks, D. L., Whittaker, M. N., Said, H., Dwivedi, G., Qu, P., Musunuru, K., Ahrens-Nicklas, R. C., Alameh, M.-G., & Wang, X. (2024). A base editing strategy using mRNA-LNPs for *in vivo* correction of the most frequent phenylketonuria variant. *Human Genetics and Genomics Advances*, 5(1), 100253. <https://doi.org/10.1016/j.xhgg.2023.100253>
- Burgard, P., Ballhausen, D., Hennermann, J. B., Hoffmann, G. F., Kölker, S., Konstantopoulou, V., Lachmann, R., Maier, E. M., Murphy, E., Ullrich, K., Ziaqaki, A., Zschocke, J., & Lindner, M. (2026). Correspondence on “Phenylalanine hydroxylase deficiency diagnosis and management: A 2023 evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG)” by Smith *et al.* *Genetics in Medicine*, 28(1), 101600. <https://doi.org/10.1016/j.gim.2025.101600>

- Burlina, A. P., Cazzorla, C., Massa, P., Loro, C., Gueraldi, D., & Burlina, A. B. (2020). The impact of a slow-release large neutral amino acids supplement on treatment adherence in adult patients with phenylketonuria. *Nutrients*, 12(7), 2078. <https://doi.org/10.3390/nu12072078>
- Burton, B. K., Clague, G. E., Harding, C. O., Kucuksayrac, E., Levy, D. G., Lindstrom, K., Longo, N., Maillot, F., Muntau, A. C., Rutsch, F., & Zori, R. T. (2024). Long-term comparative effectiveness of pegvaliase versus medical nutrition therapy with and without sapropterin in adults with phenylketonuria. *Molecular Genetics and Metabolism*, 141(1), 108114. <https://doi.org/10.1016/j.ymgme.2023.108114>
- Burton, B. K., Hermida, Á., Bélanger-Quintana, A., Bell, H., Bjoraker, K. J., Christ, S. E., Grant, M. L., Harding, C. O., Huijbregts, S. C. J., Longo, N., McNutt, M. C., Nguyen-Driver, M. D., Santos Pessoa, A. L., Rocha, J. C., Sacharow, S., Sanchez-Valle, A., Sivri, H. S., Vockley, J., Walterfang, M., . . . Muntau, A. C. (2022). Management of early treated adolescents and young adults with phenylketonuria: Development of international consensus recommendations using a modified Delphi approach. *Molecular Genetics and Metabolism*, 137(1–2), 114–126. <https://doi.org/10.1016/j.ymgme.2022.07.012>
- Charbonneau, M. R., Denney, W. S., Horvath, N. G., Cantarella, P., Castillo, M. J., Puurunen, M. K., & Brennan, A. M. (2021). Development of a mechanistic model to predict synthetic biotic activity in healthy volunteers and patients with phenylketonuria. *Communications Biology*, 4(1), 898. <https://doi.org/10.1038/s42003-021-02183-1>
- Diaz-Trelles, R., Lee, S., Kuakini, K., Park, J., Dukanovic, A., Gonzalez, J. A., Dam, T., Kim, J. H., Vega, J. B., Sablad, M., Karmali, P. P., Tachikawa, K., & Chivukula, P. (2022). Lipid nanoparticle delivers phenylalanine ammonia lyase mRNA to the liver leading to catabolism and clearance of phenylalanine in a phenylketonuria mouse model. *Molecular Genetics and Metabolism Reports*, 32, 100882. <https://doi.org/10.1016/j.ymgmr.2022.100882>
- Evans, S., Singh, R., Ahring, K., Ashmore, C., Daly, A., Ford, S., Gama, M. I., Giżewska, M., Hill, M., Ilgaz, F., Jackson, R., Newby, C., Pinto, A., Tosi, M., Nas, O. Y., Zuvadelli, J., & MacDonald, A. (2026). Global use of casein glycomacropeptide protein substitutes for phenylketonuria (PKU): Health professional perspectives. *Nutrients*, 18(3), 488. <https://doi.org/10.3390/nu18030488>
- Feillet, F., Arnoux, J.-B., Delgado, M. B., Burlina, A., Chabrol, B., Kucuksayrac, E., Lagler, F. B., Muntau, A. C., Olsson, D., Paci, S., Rutsch, F., & van Spronsen, F. J. (2025). Long-term safety of sapropterin in paediatric and adult individuals with phenylalanine hydroxylase deficiency: Final results of the Kuvan® adult maternal paediatric European registry multinational observational study. *Journal of Inherited Metabolic Disease*, 48(1), e12796. <https://doi.org/10.1002/jimd.12796>
- Feillet, F., Ficicioglu, C., Lagler, F. B., Longo, N., Muntau, A. C., Burlina, A., Trefz, F. K., van Spronsen, F. J., Arnoux, J.-B., Lindstrom, K., Lilienstein, J., Clague, G. E., Rowell, R., & Burton, B. K. (2024). Efficacy and safety of sapropterin before and during pregnancy: Final analysis of the Kuvan® adult maternal paediatric European registry (KAMPER) maternal and phenylketonuria developmental outcomes and safety (PKUDOS) PKU-MOMs sub-registries. *Journal of Inherited Metabolic Disease*, 47(4), 636–650. <https://doi.org/10.1002/jimd.12724>
- Fitzpatrick, P. F. (2023). The aromatic amino acid hydroxylases: Structures, catalysis, and regulation of phenylalanine hydroxylase, tyrosine hydroxylase, and tryptophan hydroxylase. *Archives of Biochemistry and Biophysics*, 735, 109518. <https://doi.org/10.1016/j.abb.2023.109518>
- Garbade, S. F., Shen, N., Himmelreich, N., Haas, D., Trefz, F. K., Hoffmann, G. F., Burgard, P., & Blau, N. (2019). Allelic phenotype values: A model for genotype-based phenotype prediction in phenylketonuria. *Genetics in Medicine*, 21(3), 580–590. <https://doi.org/10.1038/s41436-018-0081-x>
- Gautier, M. K., King, K., Han, Y., Said, H., Alameh, M.-G., Wang, X., Hong, X., Musunuru, K., & Ahrens-Nicklas, R. C. (2026). *In vivo* adenine base editing rescues brain biochemistry and improves motor deficits in a common phenylketonuria variant. *Molecular Therapy Advances*, 34(2), 201760. <https://doi.org/10.1016/j.omta.2026.201760>
- Giżewska, M., Inwood, A., Tyčová, R., Vijay, S., Fjellbirkeland, O., van Spronsen, F., Venegas-Moreno, E. M., Guilder, L., Burlina, A., Peters, H., Potter, M., Grošelj, U., Chakrapani, A., Bélanger-Quintana, A., Maillot, F., Rutsch, F., Arnoux, J.-B., Tchan, M., Ingalls, K., . . . Muntau, A. C. (2026). Efficacy and safety of sepiapterin versus sapropterin in patients with phenylketonuria: Results from the phase 3, randomized, crossover, open-label, active-controlled AMPLIPHY trial. *Metabolism*, 178, 156513. <https://doi.org/10.1016/j.metabol.2026.156513>
- Gundorova, P., Yousefi, B., Woidy, M., Rose-Heine, M. S., Khatri, R., Kasten, V., Bonn, S., Muntau, A. C., & Gersting, S. W. (2025). Personalized genotype-based approach for treatment of phenylketonuria. *Journal of Inherited Metabolic Disease*, 48(5), e70067. <https://doi.org/10.1002/jimd.70067>
- Haitjema, S., van Steenis, E. M., Evers, R. A. F., Boer, T., Lubout, C. M. A., van Spronsen, F. J., & Heiner-Fokkema, M. R. (2026). Toward personalized treatment in phenylketonuria: Intra-individual variation of *in vivo* phenylalanine oxidation using the 13C-phenylalanine breath test. *Molecular Genetics and Metabolism*, 148(4), 110183. <https://doi.org/10.1016/j.ymgme.2026.110183>
- Harding, C. O., Amato, R. S., Stuy, M., Longo, N., Burton, B. K., Posner, J., Weng, H. H., Merilainen, M., Gu, Z., Jiang, J., & Vockley, J. (2018). Pegvaliase for the treatment of phenylketonuria: A pivotal, double-blind randomized discontinuation phase 3 clinical trial. *Molecular Genetics and Metabolism*, 124(1), 20–26. <https://doi.org/10.1016/j.ymgme.2018.03.003>
- Harding, C. O., Longo, N., Northrup, H., Sacharow, S., Singh, R., Thomas, J. A., Vockley, J., Zori, R. T., Bulloch Whitehall, K., Lilienstein, J., Lindstrom, K., Levy, D. G., Jones, S., & Burton, B. K. (2024). Pegvaliase for the treatment of phenylketonuria: Final results of a long-term phase 3 clinical trial program. *Molecular Genetics and Metabolism Reports*, 39, 101084. <https://doi.org/10.1016/j.ymgmr.2024.101084>
- Hillert, A., Anikster, Y., Belanger-Quintana, A., Burlina, A., Burton, B. K., Carducci, C., Chiesa, A. E., Christodoulou, J., Đorđević, M., Desviat, L. R., Eliyahu, A., Evers, R. A. F., Fajkusova, L., Feillet, F., Bonfim-Freitas, P. E., Giżewska, M., Gundorova, P., Karall, D., Kneller, K., . . . Blau, N. (2020). The genetic landscape and epidemiology of phenylketonuria. *The American Journal of Human Genetics*, 107(2), 234–250. <https://doi.org/10.1016/j.ajhg.2020.06.006>

- Hollander, S., Valli, B., Vucko, E., Lah, M., Sanchez-Valle, A., Murray, B. M., Kritzer, A., & Sacharow, S. (2026). Pegvalias treatment for adolescents with phenylketonuria: A multi-site study. *JIMD Reports*, 67(2), e70070. <https://doi.org/10.1002/jmd2.70070>
- Jones, H., & Strehle, E.-M. (2026). Nutritional and therapeutic strategies in paediatric phenylketonuria: A narrative literature review. *Nutrients*, 18(9), 1347. <https://doi.org/10.3390/nu18091347>
- Kalter, N., Fuster-García, C., Silva, A., Ronco-Díaz, V., Roncelli, S., Turchiano, G., Gorodkin, J., Cathomen, T., Benabdellah, K., Lee, C., & Hendel, A. (2025). Off-target effects in CRISPR-Cas genome editing for human therapeutics: Progress and challenges. *Molecular Therapy Nucleic Acids*, 36(3), 102636. <https://doi.org/10.1016/j.omtn.2025.102636>
- Kang, M. J. Y., Adamson, C., Malpas, C. B., Winton-Brown, T., Burgess, N., Fazio, T., Velakoulis, D., Desmond, P., Venkatraman, V., De Jong, G., & Walterfang, M. (2026). Improved dietary control results in improvements in indices of white matter structure in adults with phenylketonuria: The ReDAPT study. *Acta Neuropsychiatrica*, 38, e39. <https://doi.org/10.1017/neu.2026.10078>
- Karaca, M., Gedikbasi, A., Selamioglu, A., Yesil, A., Toksoy, G., Gulec, C., Balci, M. C., Güneş, D., Yıldırım, B. T., Kına, G., Karaman, V., Aslanger, A. D., Karaman, B., Demirkol, M., Uyguner, Z. O., & Gokcay, G. F. (2026). Genotype–phenotype correlations in phenylketonuria: PAH variants and BH4 responsiveness for treatment design. *Human Genomics*, 20(1), 70. <https://doi.org/10.1186/s40246-026-00936-9>
- Kumar, R. K., & Prabha, N. (2026). Eighteen years of experience with newborn screening in India. *Indian Journal of Pediatrics*, 93(6), 645. <https://doi.org/10.1007/s12098-026-06141-9>
- Lamb, Y. N. (2026). Sepiapterin: First approval. *Drugs*, 86(1), 93–99. <https://doi.org/10.1007/s40265-025-02247-0>
- Langeveld, M., Sirrs, S., Lachman, R., Fazio, T., Sharma, R., Ziagaki, A., Reismann, P., Tchan, M., Tran, C., van der Klauw, M. M., Vassiliou, D., & Murphy, E. (2026). Treatment goals for adults with early treated PKU should be determined by evidence-based shared decision making between patients and their medical team. *Orphanet Journal of Rare Diseases*, 21(1), 233. <https://doi.org/10.1186/s13023-026-04376-5>
- Longo, N., Burton, B. K., Vaghela, S., Mu, F., Wang, A., Smith, J., Reichert, E., Simpson, R. B., Sikirica, V., & Perera, S. (2026). Disease burden and pharmacological treatment patterns in children and adults with phenylketonuria: A real-world matched cohort study. *Journal of Inherited Metabolic Disease*, 49(3), e70194. <https://doi.org/10.1002/jimd.70194>
- Longo, N., Dimmock, D., Levy, H., Viau, K., Bausell, H., Bilder, D. A., Burton, B., Gross, C., Northrup, H., Rohr, F., Sacharow, S., Sanchez-Valle, A., Stuy, M., Thomas, J., Vockley, J., Zori, R., & Harding, C. O. (2019). Evidence- and consensus-based recommendations for the use of pegvalias in adults with phenylketonuria. *Genetics in Medicine*, 21(8), 1851–1867. <https://doi.org/10.1038/s41436-018-0403-z>
- Longo, N., Hamazaki, T., Hollander, S., MacDonald, A., Muntau, A. C., & Schwartz, I. V. D. (2025). Global considerations for lifelong management and therapeutic development for phenylketonuria. *Genetics in Medicine*, 27(11), 101540. <https://doi.org/10.1016/j.gim.2025.101540>
- MacDonald, A., Ahring, K., Bledsoe, A., Fujimoto, H., Giorda, S., Kogelmann, C., Kopesky, J., Nagy, L., O'Neill, S., Pinto, A., Poloni, S., Roberts, P., van Wegberg, A. M. J., & Hollander, S. (2026). Dietetic guidance for nutritional management of people with phenylketonuria receiving sepiapterin. *Molecular Genetics and Metabolism*, 147(1), 109705. <https://doi.org/10.1016/j.ymgme.2025.109705>
- Martinez, M., Harding, C. O., Schwank, G., & Thöny, B. (2024). State-of-the-art 2023 on gene therapy for phenylketonuria. *Journal of Inherited Metabolic Disease*, 47(1), 80–92. <https://doi.org/10.1002/jimd.12651>
- Merkel, M., Berg, D., Brüggemann, N., Classen, J., Mainka, T., Zittel, S., & Muntau, A. C. (2023). Characterisation and differential diagnosis of neurological complications in adults with phenylketonuria: Literature review and expert opinion. *Journal of Neurology*, 270(8), 3675–3687. <https://doi.org/10.1007/s00415-023-11703-4>
- Muntau, A. C., Adams, D. J., Bélanger-Quintana, A., Bushueva, T. V., Cerone, R., Chien, Y.-H., Chiesa, A., Coşkun, T., de las Heras, J., Feillet, F., Katz, R., Lagler, F., Piazzon, F., Rohr, F., van Spronsen, F. J., Vargas, P., Wilcox, G., & Bhattacharya, K. (2019). International best practice for the evaluation of responsiveness to sapropterin dihydrochloride in patients with phenylketonuria. *Molecular Genetics and Metabolism*, 127(1), 1–11. <https://doi.org/10.1016/j.ymgme.2019.04.004>
- Muntau, A. C., Longo, N., Ezgu, F., Schwartz, I. V. D., Lah, M., Bratkovic, D., Margvelashvili, L., Kiykim, E., Zori, R., Campistol Plana, J., Bélanger-Quintana, A., Lund, A., Guider, L., Chakrapani, A., Mungan, H. N., Guimas, A., Cabrales Guerra, I. D. C., MacDonald, A., Ingalls, K., . . . Smith, N. (2024). Effects of oral sepiapterin on blood Phe concentration in a broad range of patients with phenylketonuria (APHENITY): Results of an international, phase 3, randomised, double-blind, placebo-controlled trial. *The Lancet*, 404(10460), 1333–1345. [https://doi.org/10.1016/s0140-6736\(24\)01556-3](https://doi.org/10.1016/s0140-6736(24)01556-3)
- Muntau, A. C., Zhang, R., Chakrapani, A., Hamazaki, T., Lah, M., Ruebel, D. J., Vijay, S., Zori, R. T., Feillet, F., Longo, N., Northrup, H., Hollander, S., O'Connell, T., Woolley, J. J., & Tomazos, I. (2026). Development of a face-validated conceptual model structure for economic evaluation in phenylketonuria, through expert elicitation. *Orphanet Journal of Rare Diseases*, 21(1), 262. <https://doi.org/10.1186/s13023-026-04423-1>
- Ney, D. M., Stroup, B. M., Clayton, M. K., Murali, S. G., Rice, G. M., Rohr, F., & Levy, H. L. (2016). Glycomacropeptide for nutritional management of phenylketonuria: A randomized, controlled, crossover trial. *The American Journal of Clinical Nutrition*, 104(2), 334–345. <https://doi.org/10.3945/ajcn.116.135293>
- Nicosia, L. (2026). Balancing personalized medicine and scalability in base editing for phenylketonuria. *Molecular Therapy Nucleic Acids*, 37(1), 102824. <https://doi.org/10.1016/j.omtn.2025.102824>
- Nulmans, I., Lequeue, S., Desmet, L., Del'hay, G. G., Salvi, N. S., Rombaut, M., Neuckermans, J., & De Kock, J. (2025). Phenylketonuria: A guide through the complex maze of its neurological pathophysiology providing a new perspective on treatment strategies. *Biomedicine & Pharmacotherapy*, 191, 118522. <https://doi.org/10.1016/j.biopha.2025.118522>
- Peres, M., Almeida, M. F., Pinto, É. J., Carmona, C., Rocha, S., Guimas, A., Ribeiro, R., Martins, E., Bandeira, A., MacDonald, A., & Rocha, J. C. (2021). Implementing a transition program from paediatric to adult services in phenylketonuria: Results after two years of follow-up with an adult team. *Nutrients*, 13(3), 799. <https://doi.org/10.3390/nu13030799>

- Perez-Garcia, C. G., Diaz-Trelles, R., Vega, J. B., Bao, Y., Sablad, M., Limphong, P., Chikamatsu, S., Yu, H., Taylor, W., Karmali, P. P., Tachikawa, K., & Chivukula, P. (2022). Development of an mRNA replacement therapy for phenylketonuria. *Molecular Therapy Nucleic Acids*, 28, 87–98. <https://doi.org/10.1016/j.omtn.2022.02.020>
- Pey, A. L., Ying, M., Cremades, N., Velazquez-Campoy, A., Scherer, T., Thöny, B., Sancho, J., & Martinez, A. (2008). Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. *Journal of Clinical Investigation*, 118(8), 2858–2867. <https://doi.org/10.1172/jci34355>
- Pinto, A., Ahring, K., Almeida, M. F., Ashmore, C., Bailey, S., Bélanger-Quintana, A., Burlina, A., Cole, D., Dale, C., Daly, A., van Dam, E., Dawson, C., Evans, S., Firman, S., Ford, S., Green, D., Geberhiwot, T., Hoekstra, Y., Howe, S., . . . MacDonald, A. (2026). How well is blood phenylalanine controlled in maternal PKU in Europe? Results from 102 pregnancies. *Nutrients*, 18(13), 2136. <https://doi.org/10.3390/nu18132136>
- Pinto, A., Daly, A., Newby, C., Robotham, A., Heales, S., Eaton, S., Aitkenhead, H., Gilmour, K., Jackson, R., Ashmore, C., Evans, S., Rocha, J. C., Ilgaz, F., Hickson, M., & MacDonald, A. (2024). The effects of casein glycomacropeptide on general health status in children with PKU: A randomized crossover trial. *Molecular Genetics and Metabolism*, 143(4), 108607. <https://doi.org/10.1016/j.ymgme.2024.108607>
- Quigley, A., Jindal, I., Campion, T., Rutherford, D., Han, Y., Quigley, W., Qu, P., Musunuru, K., Ahrens-Nicklas, R., Hong, X., & Wang, X. (2025). Base editing strategies for *in vivo* correction of two highly recurrent phenylketonuria variants. *Molecular Therapy Nucleic Acids*, 36(4), 102770. <https://doi.org/10.1016/j.omtn.2025.102770>
- Remor, E., Gabe, K. M., Teruya, K. I., & Doederlein Schwartz, I. V. (2024). What is known about patients' quality of life with phenylketonuria and their caregivers? A scoping review. *Orphanet Journal of Rare Diseases*, 19(1), 402. <https://doi.org/10.1186/s13023-024-03422-4>
- Rocha, J. C., & Martel, F. (2009). Large neutral amino acids supplementation in phenylketonuric patients. *Journal of Inherited Metabolic Disease*, 32(4), 472–480. <https://doi.org/10.1007/s10545-009-1132-x>
- Rocha, J. C., Hermida, Á., Jones, C. J., Wu, Y., Clague, G. E., Rose, S., Whitehall, K. B., Ahring, K. K., Pessoa, A. L. S., Harding, C. O., Rohr, F., Inwood, A., Longo, N., Muntau, A. C., Sivri, S., & Maillot, F. (2024). Meta-analysis of bone mineral density in adults with phenylketonuria. *Orphanet Journal of Rare Diseases*, 19(1), 338. <https://doi.org/10.1186/s13023-024-03223-9>
- Rohde, C., Thiele, A. G., Baerwald, C., Ascherl, R. G., Lier, D., Och, U., Heller, C., Jung, A., Schönherr, K., Joerg-Streller, M., Luttat, S., Matzgen, S., Winkler, T., Rosenbaum-Fabian, S., Joos, O., & Beblo, S. (2021). Preventing maternal phenylketonuria (PKU) syndrome: Important factors to achieve good metabolic control throughout pregnancy. *Orphanet Journal of Rare Diseases*, 16(1), 477. <https://doi.org/10.1186/s13023-021-02108-5>
- Romani, C., Huijbregts, S., van Spronsen, F. J., MacDonald, A., van Wegberg, A. M. J., Staines, M., & Olson, A. (2025). Meta-analysis of cognitive outcomes in children and adults with early treated phenylketonuria – Results across functions. *Molecular Genetics and Metabolism*, 146(1–2), 109210. <https://doi.org/10.1016/j.ymgme.2025.109210>
- Romani, C., Olson, A., Aitkenhead, L., Baker, L., Patel, D., van Spronsen, F. J., MacDonald, A., Wegberg, A. V., & Huijbregts, S. (2022). Meta-analyses of cognitive functions in early-treated adults with phenylketonuria. *Neuroscience & Biobehavioral Reviews*, 143, 104925. <https://doi.org/10.1016/j.neubiorev.2022.104925>
- Rothgangl, T., Tálas, A., Ioannidi, E. I., Weber, Y., Böck, D., Matsushita, M., Villiger, E. A., Schmidheini, L., Moon, W. J., Lin, P. J. C., Fan, S. H. Y., Marquart, K. F., Schwerdel, C., Rimann, N., Faccin, E., Villiger, L., Muramatsu, H., Vadovics, M., Cremonesi, A., . . . Schwank, G. (2025). Treatment of a metabolic liver disease in mice with a transient prime editing approach. *Nature Biomedical Engineering*, 9(10), 1705–1718. <https://doi.org/10.1038/s41551-025-01399-4>
- Santos-Sierra, S., Kirchmair, J., Perna, A. M., Reiß, D., Kemter, K., Röschinger, W., Glossmann, H., Gersting, S. W., Muntau, A. C., Wolber, G., & Lagler, F. B. (2012). Novel pharmacological chaperones that correct phenylketonuria in mice. *Human Molecular Genetics*, 21(8), 1877–1887. <https://doi.org/10.1093/hmg/dds001>
- Scala, I., Riccio, M. P., Marino, M., Bravaccio, C., Parenti, G., & Strisciuglio, P. (2020). Large neutral amino acids (LNAAs) supplementation improves neuropsychological performances in adult patients with phenylketonuria. *Nutrients*, 12(4), 1092. <https://doi.org/10.3390/nu12041092>
- Smith, W. E., Berry, S. A., Bloom, K., Brown, C., Burton, B. K., Demarest, O. M., Jenkins, G. P., Malinowski, J., McBride, K. L., Mroczkowski, H. J., Scharfe, C., & Vockley, J. (2025). Phenylalanine hydroxylase deficiency diagnosis and management: A 2023 evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 27(1), 101289. <https://doi.org/10.1016/j.gim.2024.101289>
- Thomas, J., Levy, H., Amato, S., Vockley, J., Zori, R., Dimmock, D., Harding, C. O., Bilder, D. A., Weng, H. H., Olbertz, J., Merilainen, M., Jiang, J., Larimore, K., Gupta, S., Gu, Z., & Northrup, H. (2018). Pegvaliase for the treatment of phenylketonuria: Results of a long-term phase 3 clinical trial program (PRISM). *Molecular Genetics and Metabolism*, 124(1), 27–38. <https://doi.org/10.1016/j.ymgme.2018.03.006>
- Trampuž, D., Schielen, P. C. J. I., Zetterström, R. H., Scarpa, M., Feillet, F., Kožich, V., Tangeras, T., Drole Torkar, A., Mlinarič, M., Perko, D., Remec, Ž. I., Lampret, B. R., Battelino, T., van Spronsen, F. J., Bonham, J. R., & Grošelj, U. (2025). International survey on phenylketonuria newborn screening. *International Journal of Neonatal Screening*, 11(1), 18. <https://doi.org/10.3390/ijns11010018>
- Trepp, R., Muri, R., Maissen-Abgottsporn, S., Haynes, A. G., Hochuli, M., & Everts, R. (2024). Cognition after a 4-week high phenylalanine intake in adults with phenylketonuria – a randomized controlled trial. *The American Journal of Clinical Nutrition*, 119(4), 908–916. <https://doi.org/10.1016/j.ajcnut.2023.11.007>
- Turki, A., Stockler-Ipsiroglu, S., Sirrs, S., Branov, J., Bosdet, T., & Elango, R. (2025). Protein requirements in adults with phenylketonuria and bioavailability of glycomacropeptide compared to an L-amino acid-based product. *Journal of Inherited Metabolic Disease*, 48(1), e12806. <https://doi.org/10.1002/jimd.12806>
- van Spronsen, F. J., Blau, N., Harding, C., Burlina, A., Longo, N., & Bosch, A. M. (2021). Phenylketonuria. *Nature Reviews Disease Primers*, 7(1), 36. <https://doi.org/10.1038/s41572-021-00267-0>

- van Spronsen, F. J., van Wegberg, A. M., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., Burlina, A., Campistol, J., Feillet, F., Giżewska, M., Huijbregts, S. C., Kearney, S., Leuzzi, V., Mailliot, F., Muntau, A. C., Trefz, F. K., van Rijn, M., Walter, J. H., & MacDonald, A. (2017). Key European guidelines for the diagnosis and management of patients with phenylketonuria. *The Lancet Diabetes & Endocrinology*, 5(9), 743–756. [https://doi.org/10.1016/s2213-8587\(16\)30320-5](https://doi.org/10.1016/s2213-8587(16)30320-5)
- van Spronsen, F., Peters, H., Margvelashvili, L., Agladze, D., Schwartz, I. V. D., Giżewska, M., Hamazaki, T., Guilder, L., MacDonald, A., Vijay, S., Inwood, A., Minami, M., Fjellbirkeland, O., Lund, A., Lah, M., Thomas, J. A., Longo, N., Kiykim, E., Ishige, M., . . . Muntau, A. C. (2026). Effect of long-term sepiapterin treatment on dietary phenylalanine tolerance in patients with phenylketonuria: Interim results from the phase 3 APHENITY extension study. *Genetics in Medicine*, 28(4), 101683. <https://doi.org/10.1016/j.gim.2026.101683>
- van Wegberg, A. M. J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A. M., Burlina, A., Campistol, J., Feillet, F., Giżewska, M., Huijbregts, S. C., Kearney, S., Leuzzi, V., Mailliot, F., Muntau, A. C., van Rijn, M., Trefz, F., Walter, J. H., & van Spronsen, F. J. (2017). The complete European guidelines on phenylketonuria: Diagnosis and treatment. *Orphanet Journal of Rare Diseases*, 12(1), 162. <https://doi.org/10.1186/s13023-017-0685-2>
- van Wegberg, A. M. J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Beblo, S., Blau, N., Bosch, A. M., Burlina, A., Campistol, J., Coşkun, T., Feillet, F., Giżewska, M., Huijbregts, S. C., Leuzzi, V., Mailliot, F., Muntau, A. C., Rocha, J. C., Romani, C., Trefz, F., & van Spronsen, F. J. (2025). European guidelines on diagnosis and treatment of phenylketonuria: First revision. *Molecular Genetics and Metabolism*, 145(2), 109125. <https://doi.org/10.1016/j.ymgme.2025.109125>
- Vockley, J., Sondheimer, N., Puurunen, M., Diaz, G. A., Ginevic, I., Grange, D. K., Harding, C., Northrup, H., Phillips, J. A., Searle, S., Thomas, J. A., Zori, R., Denney, W. S., Ernst, S. L., Humphreys, K., McWhorter, N., Kurtz, C., & Brennan, A. M. (2023). Efficacy and safety of a synthetic biotic for treatment of phenylketonuria: A phase 2 clinical trial. *Nature Metabolism*, 5(10), 1685–1690. <https://doi.org/10.1038/s42255-023-00897-6>
- Whitehall, K. B., Rose, S., Clague, G. E., Ahring, K. K., Bilder, D. A., Harding, C. O., Hermida, Á., Inwood, A., Longo, N., Mailliot, F., Muntau, A. C., Pessoa, A. L. S., Rocha, J. C., Rohr, F., Sivri, S., Said, J., Oshinbolu, S., & Sibbring, G. C. (2024). Systematic literature review of the somatic comorbidities experienced by adults with phenylketonuria. *Orphanet Journal of Rare Diseases*, 19(1), 293. <https://doi.org/10.1186/s13023-024-03203-z>
- Whittaker, M. N., Testa, L. C., Quigley, A., Brooks, D. L., Grandinette, S. A., Said, H., Dwivedi, G., Jindal, I., Volpp, D., Hacker, J. L., Qu, P., Wang, J. Z., Levine, M. A., Ahrens-Nicklas, R. C., Li, Q., Musunuru, K., Alameh, M.-G., Peranteau, W. H., & Wang, X. (2026). Improved specificity and efficiency of *in vivo* adenine base editing therapies with hybrid guide RNAs. *Nature Biomedical Engineering*, 10(9), 1943–1955. <https://doi.org/10.1038/s41551-025-01545-y>
- Woolf, L. I., & Adams, J. (2020). The early history of PKU. *International Journal of Neonatal Screening*, 6(3), 59. <https://doi.org/10.3390/ijns6030059>
- Yin, S., Gao, L., Sun, X., Zhang, M., Gao, H., Chen, X., Zhang, D., Ming, X., Yang, L., Hu, Y., Chen, X., Liu, M., Zhan, X., Guan, Y., Wang, L., Han, L., Zhu, P., & Li, D. (2025). Amelioration of metabolic and behavioral defects through base editing in the Pah R408W phenylketonuria mouse model. *Molecular Therapy*, 33(1), 119–132. <https://doi.org/10.1016/j.ymthe.2024.11.032>
- Yu, I., & Jeong, J. (2025). Advancing gene therapy for phenylketonuria: From precision editing to clinical translation. *International Journal of Molecular Sciences*, 26(17), 8722. <https://doi.org/10.3390/ijms26178722>

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2026): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

#### Peer-review history:

The peer review history for this paper can be accessed here: <https://pr.sdiarticle5.com/review-history/167058>