Asian Journal of Medicine and Health

Volume 24, Issue 9, Page 230-252, 2026; Article no.AJMAH.164501
ISSN: 2456-8414

Artificial Intelligence in the Analysis of the Fetal Genome in Utero: A Critical Review of Current Paradigms, Clinical Utility and Future Horizons

S. Bittmann [a*], E. Luchter [a] and E. Moschüring-Alieva [a]

[a] Department of Pediatrics, Ped Mind Institute, Hindenburgring 4, 48599 Gronau, Germany.

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/ajmah/2026/v24i91430

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/164501

Review Article

Received: 30/06/2026
Accepted: 26/08/2026
Published: 29/08/2026

Abstract

Prenatal genomic medicine now generates data at a scale that exceeds what unaided human interpretation can process within the time available to a continuing pregnancy. Cell-free DNA screening, chromosomal microarray, exome sequencing and, increasingly, genome sequencing produce millions of candidate observations per case, and computational methods described as artificial intelligence have been proposed at almost every step of the resulting pipeline. This critical narrative review evaluates the strength, consistency and clinical relevance of the evidence supporting those methods when they are applied specifically to the genome of a fetus in utero. Literature was identified through structured searching of biomedical and multidisciplinary scholarly sources from 1997 to 20 June 2026, prioritised by methodological quality and by direct relevance to prenatal application rather than by citation count. The evidence divides sharply. Analytical performance for tasks that are essentially signal-detection problems, including trisomy classification from low-coverage plasma sequencing, fetal fraction estimation and small-variant calling, is supported by internally consistent and technically credible work. Evidence for tasks that require inference about meaning, including missense and splice-effect prediction, phenotype-driven gene prioritisation and automated candidate diagnosis, is substantially weaker in the prenatal context, because the models were trained and calibrated on postnatal cohorts whose phenotypes were fully observable. Fetal phenotypes evolve during gestation, are captured through imaging rather than

**Corresponding author: E-mail: stefanbittmann@gmx.de;*

Cite as: *Bittmann, S., Luchter, E., & Moschüring-Alieva, E. (2026). Artificial Intelligence in the Analysis of the Fetal Genome in Utero: A Critical Review of Current Paradigms, Clinical Utility and Future Horizons. Asian Journal of Medicine and Health, 24(9), 230–252. https://doi.org/10.9734/ajmah/2026/v24i91430*

examination, and map imperfectly onto the ontologies that phenotype-driven algorithms consume. Several studies show that ontology-driven gene selection in fetuses misses a substantial minority of causative genes that curated prenatal panels retain. Almost no prospective study has tested whether an artificial intelligence component changes a prenatal decision, and reporting quality across the field falls short of contemporary prediction-model standards. Structural inequities in reference genomic resources compound these problems. Priorities include prenatally calibrated variant-effect evidence, gestational-age-aware phenotype representation, prospective evaluation with decision-relevant endpoints, and governance that treats uncertainty as a reportable output rather than a defect to be smoothed away.

Keywords: Prenatal diagnosis; fetal genome; artificial intelligence; machine learning; cell-free DNA; variant interpretation; clinical utility.

1. Introduction

Prenatal assessment of the fetal genome has been transformed twice within three decades. The first transformation began with the demonstration that fetal DNA circulates in maternal plasma and serum (Lo *et al*., 1997), which established a route to fetal genetic information that did not require entry into the uterus. Within little more than a decade, sequencing of maternal plasma had been used to reconstruct the genome-wide genetic and mutational profile of the fetus (Lo *et al*., 2010) and to assemble a noninvasive whole-genome sequence of a human fetus from parental haplotypes and plasma allele counts (Kitzman *et al*., 2012). Screening for the common autosomal trisomies using circulating cell-free DNA subsequently achieved detection rates and false-positive rates that no serum-and-ultrasound algorithm had matched (Gil *et al*., 2017), and professional bodies moved from cautious endorsement in high-risk pregnancies to evidence-based recommendation across the general-risk population (Dungan *et al*., 2023).

Recent work extending cell-free plasma DNA analysis to genome-wide haplotype-based diagnosis of monogenic disorders further illustrates the expanding analytical scope of noninvasive fetal genomic assessment (Zhang *et al*., 2026).

The second transformation concerned diagnostic rather than screening technology. Two prospective cohort studies published simultaneously established that exome sequencing of fetuses with structural anomalies and normal karyotype and microarray results yields diagnostic variants in a clinically meaningful minority of cases: 8.5% of 610 fetuses in the Prenatal Assessment of Genomes and Exomes cohort (Lord *et al*., 2019) and 10% of 234 trios in an unselected North American cohort (Petrovski *et al*., 2019). Genome sequencing has since been evaluated against microarray and exome pathways, with a pooled incremental yield over quantitative fluorescence polymerase chain reaction and chromosomal microarray analysis of 16% for prenatal cases (Shreeve *et al*., 2024), and prospective multicentre data now exist for genome sequencing applied directly to fetal structural anomalies (Gao *et al*., 2026). National services have been built around these findings, with formal evaluations of implementation (Walton *et al*., 2025) and cost-effectiveness (Smith *et al*., 2025).

Recent prenatal exome studies also indicate that diagnostic performance remains strongly dependent on the clinical indication and the quality of fetal phenotyping, reinforcing the need to interpret sequencing results within the specific prenatal phenotype (El-Dessouky *et al*., 2025; Zemet *et al*., 2025).

Both transformations share a computational consequence. A single low-coverage plasma sequencing run, a trio exome or a fetal genome produces a volume of candidate observations that cannot be adjudicated exhaustively by a human analyst within the timescale that matters clinically, which in a continuing pregnancy may be a small number of days. Methods described under the heading of artificial intelligence have accordingly been proposed at almost every step: classification of aneuploidy from fragment-level features of plasma DNA (Lee *et al*., 2022), estimation of the fetal fraction from fragment length and count distributions (Gazdarica *et al*., 2019), detection of fetal single-nucleotide variants in plasma using self-learning networks (Qi *et al*., 2024), pan-genomic modelling to reduce reference bias in noninvasive testing (Xue *et al*., 2024), deep neural variant calling from short-read alignments (Poplin *et al*., 2018), prediction of splicing consequences from primary sequence (Jaganathan *et al*., 2019), proteome-wide missense effect prediction (Cheng *et al*., 2023), phenotype-driven ranking of candidate genes (Cooperstein *et al*., 2025), and integrated systems that nominate candidate diagnoses from a genome and a phenotype description (Clark *et al*., 2019; De La Vega *et al*., 2021).

The difficulty is that these methods were, with very few exceptions, developed and validated in settings that differ from the prenatal setting in ways that bear directly on their behaviour. Postnatal rare-disease cohorts supply a phenotype that has been examined, described and, where necessary, re-examined. A fetus supplies a phenotype inferred from ultrasound and magnetic resonance imaging, observed at one or two points in a developmental trajectory, and liable to change. A cohort study of consecutive prenatal exome cases found that additional anomalies emerged between the second and third trimesters in 73.3% of cases that reached the third trimester with a causative variant identified, and that variants of uncertain significance were reclassified on the basis of postnatal information (Mone *et al*., 2022). Reanalysis of unsolved prenatal exomes using postnatal and post-mortem features has been shown to add diagnoses that the prenatal phenotype alone did not support (Thauvin-Robinet *et al*., 2025). Any algorithm whose ranking depends on the phenotype supplied to it inherits this instability. Consistent with this limitation, postnatal reanalysis of initially negative or inconclusive prenatal exomes has identified additional diagnoses when new phenotypic findings became available, underscoring the temporal dependence of sequence interpretation in fetal cases (Swanson *et al*., 2026).

A second difference concerns consequence. In postnatal rare-disease diagnostics, a delayed or missed diagnosis is usually recoverable. In the prenatal setting, results feed into decisions about the continuation of a pregnancy, about delivery planning and about fetal or immediate neonatal therapy, and those decisions are frequently irreversible and bounded by statutory gestational

limits. Professional guidance has consistently emphasised that prenatal sequencing requires careful case selection, multidisciplinary review and conservative reporting (Monaghan *et al*., 2020; Van den Veyver *et al*., 2022). Automation that improves average throughput while degrading behaviour at the extremes carries a different weight here than it does elsewhere in genomic medicine.

Existing reviews have not addressed this intersection directly. The most substantial recent syntheses of artificial intelligence in obstetrics concern imaging: a scoping review of artificial intelligence in obstetric ultrasound (Horgan *et al*., 2023) and an appraisal of deep learning for fetal ultrasound (Ramirez Zegarra & Ghi, 2023) both examine image acquisition, segmentation, biometry and anomaly detection, and neither engages with genomic interpretation. Reviews of machine learning in cell-free DNA diagnostics are dominated by oncological applications, where labelled outcome data are abundant and the tissue of origin is the object of inference (Tsui *et al*., 2025). A broader account of noninvasive prenatal testing and its genomic uses describes methodological advances without appraising the strength of the underlying evidence for algorithmic components (Jin *et al*., 2024). One review addresses the ethics of artificial intelligence in prenatal and paediatric genomic medicine and raises questions of autonomy, opacity and disability that are directly relevant, but it does not evaluate the empirical performance literature (Coghlan *et al*., 2024). A synthesis of fetal imaging, phenotyping and genomic testing comes closest to the present subject, and identifies the phenotype-genotype interface as the critical constraint, without conducting a critical appraisal of the computational tools that operate across that interface (Shear *et al*., 2025). The result is a literature in which analytical claims about algorithms and clinical claims about prenatal genomic testing have developed largely in parallel, and in which the question of whether the former improve the latter has rarely been posed as an evidential question at all.

1.1 Scope, Research Gap and Objectives

This review examines computational and statistical learning methods applied to the interrogation of the fetal genome during pregnancy. Three analytical domains are covered: cell-free DNA-based assessment of fetal chromosomal and sequence-level status from maternal plasma; sequence-level analysis of fetal DNA obtained by invasive sampling, including variant detection and variant-effect prediction; and phenotype-driven prioritisation and automated candidate-diagnosis systems as applied to fetal cases. Cross-cutting questions of evidential quality, clinical utility, equity, governance and reporting are examined in relation to these three domains.

Several adjacent fields are deliberately excluded. Algorithmic analysis of fetal ultrasound and magnetic resonance images is considered only where the imaging output functions as the phenotype input to a genomic algorithm, and not as an independent diagnostic task. Preimplantation genetic testing and polygenic embryo selection are treated only where they illuminate the governance of predictive fetal genomics, since the embryo in vitro presents a different consent structure and a different decision architecture. Maternal conditions detected incidentally through fetal screening are addressed only as a governance problem. Prediction of obstetric outcomes such as pre-eclampsia or preterm birth from maternal clinical variables is outside the scope, since the object of inference in those models is not the fetal genome.

The gap addressed is specific. The performance of learning-based methods on fetal genomic data has been reported almost entirely as analytical validity, measured against laboratory or bioinformatic reference standards. Whether such methods alter the diagnostic yield, the turnaround time, the rate of uncertain results or the decisions taken by pregnant people and clinicians has scarcely been tested, and the models most heavily relied upon for interpretation carry calibration inherited from populations in which the phenotype was fully observable. No existing review evaluates this mismatch as an integrated evidential problem spanning the plasma, the sequence and the phenotype.

The objectives are, first, to characterise the state of knowledge across the three analytical domains and to distinguish well-supported from preliminary claims; second, to explain why the fetal setting alters the behaviour and the interpretation of methods that perform acceptably elsewhere; third, to appraise the methodological quality and reporting standards of the prenatal artificial intelligence literature; fourth, to identify where evidence of clinical utility exists, where it is absent and where its absence is most consequential; and fifth, to set out prioritised research directions that would materially reduce the present uncertainty. The intended contribution is an evidence-calibrated account that separates the tasks for which computational learning is already dependable from those for which enthusiasm currently exceeds demonstration.

2. Methods for Literature Selection

2.1 Review Design and Rationale

A critical narrative design was selected because the review question spans heterogeneous evidence types that cannot be pooled meaningfully. The included material comprises algorithm-development studies reporting classification metrics, prospective and retrospective clinical cohorts reporting diagnostic yield, systematic reviews and meta-analyses of sequencing performance, professional position statements, methodological and reporting standards, and normative analyses of ethics and governance. Outcome definitions differ irreconcilably across these categories, since the area under a receiver operating characteristic curve computed on a bioinformatic benchmark and the incremental diagnostic yield of a sequencing pathway are not commensurable quantities. A systematic review restricted to a single comparable outcome would exclude most of the material bearing on the review question, and a meta-analysis would impose a spurious common metric.

The narrative design does not remove the obligation to be transparent or critical. Reporting standards for prediction models that use regression or machine learning methods were used as an appraisal reference throughout, since they specify the items that permit a

reader to judge whether a reported performance figure is likely to generalise (Collins *et al*., 2024). Studies were assessed against those items even where the authors had not adopted them, and shortfalls are reported in the synthesis rather than being treated as reasons for exclusion.

2.2 Information Sources

Searching was conducted in PubMed, Europe PMC, Crossref, OpenAlex, the Directory of Open Access Journals, Semantic Scholar and Google Scholar. Crossref was additionally used to confirm bibliographic metadata and to verify that each Digital Object Identifier resolved to the article cited. Database searching was supplemented by backward citation searching of included reviews and cohort studies, forward citation searching for influential methodological papers, related-article searching, examination of the reference lists of recent position statements, and targeted checks for corrections, expressions of concern and retractions affecting any included record. Professional society position statements and evidence-based guidelines were retrieved directly from the journals in which they were published.

2.3 Search Strategy and Search Period

Search concepts were constructed around four axes and combined with Boolean operators. The first axis captured the computational method, using terms including artificial intelligence, machine learning, deep learning, neural network, variant prioritisation and variant effect prediction. The second captured the biological substrate, using terms including fetal genome, cell-free DNA, cfDNA, fetal fraction, exome sequencing, genome sequencing and chromosomal microarray. The third captured the clinical setting, using terms including prenatal, noninvasive prenatal testing, prenatal diagnosis, fetus and in utero. The fourth captured evaluative and implementation dimensions, using terms including diagnostic yield, clinical utility, calibration, bias, equity, governance and reporting standards.

Representative strings included the following. In title-and-abstract fields: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network”) AND (fetal OR prenatal) AND (genome OR genomic OR exome OR “cell-free DNA” OR variant). In title fields for higher precision: (“noninvasive prenatal” OR “non-invasive prenatal”) AND (“deep learning” OR “machine learning” OR “neural network”). For the interpretation layer: (“variant prioritisation” OR “variant prioritization” OR “phenotype-driven” OR “Human Phenotype Ontology”) AND (prenatal OR fetal OR “rare disease”). For the clinical-evidence layer: (“prenatal exome sequencing” OR “prenatal genome sequencing”) AND (“diagnostic yield” OR “clinical utility” OR “incremental yield”). Strings were adapted to the field syntax of each source, and unrestricted keyword equivalents were used where structured field searching was unavailable.

The search period ran from 1997 to 20 June 2026. The starting year corresponds to the demonstration of fetal DNA in maternal plasma, which is the point from which any noninvasive interrogation of the fetal genome derives, and which precedes by more than a decade the first application of learned classifiers to that material. Foundational publications predating 1997 were eligible where necessary to establish conceptual or methodological lineage.

2.4 Eligibility Criteria

Eligible material comprised peer-reviewed original research, systematic reviews, meta-analyses, methodological standards and professional position statements published in indexed scholarly journals, in English, addressing the analysis of fetal genomic material during pregnancy or addressing computational methods whose outputs are used in that analysis. Studies of postnatal or paediatric genomic interpretation were eligible where the method under evaluation is applied to prenatal cases or where the study establishes the calibration properties that prenatal application inherits. Studies of algorithmic fairness, privacy, regulation and reporting were eligible where the findings bear directly on the deployment of automated genomic interpretation in clinical services.

Excluded were preprints where peer-reviewed evidence on the same question was available, conference abstracts without full methodological reporting, commercial and promotional material, non-scholarly commentary, and any record whose bibliographic identity could not be verified. Studies predicting obstetric outcomes from maternal clinical or biochemical variables without reference to fetal genomic data were excluded as outside scope. Studies of oncological cell-free DNA analysis were excluded except where the methodological development is explicitly transferable to the fetal setting and is discussed as such.

2.5 Screening, Duplicate Handling and Study Selection

Records retrieved from each source were screened by title and abstract against the eligibility criteria, and records passing that stage were assessed in full text where the full text was accessible. Duplicates arising from overlapping coverage between sources were identified by Digital Object Identifier, and where an identifier was absent by exact matching of title, first author and journal, with the peer-reviewed version retained in preference to any preprint of the same work. Where an article existed in both online-first and issue-assigned form, the issue-assigned metadata were used.

Language was restricted to English, which is a limitation acknowledged below. Records that could not be retrieved in full text were retained only where the abstract, together with verified bibliographic metadata, supported the specific claim for which the record was cited, and were otherwise excluded. No record counts are reported, because selection was purposive and iterative rather than exhaustive, and reporting counts would imply a systematic process that was not undertaken.

2.6 Critical Appraisal and Evidence Synthesis

Studies were appraised against criteria appropriate to their design. Algorithm-development studies were assessed for the independence of training and evaluation data, the provenance and representativeness of the reference standard, the presence of external validation, the reporting of calibration in addition to discrimination, the handling of class imbalance, and the transparency of the model and its inputs. Clinical cohort studies were assessed for consecutive or unselected recruitment, the completeness of prior testing, the definition and adjudication of a diagnostic result, the duration and completeness of follow-up, and the extent to which reported yield was linked to a decision or an outcome. Systematic reviews were assessed for protocol registration, the adequacy of the search, the handling of heterogeneity and the appropriateness of pooling. Position statements were treated as expert consensus rather than as empirical evidence and were cited accordingly.

Synthesis proceeded thematically. Evidence was grouped by the analytical problem being solved rather than by the technique used, since identical architectures behave differently depending on whether the target is a signal-detection problem with an objective reference standard or an inference about biological meaning with a curated and revisable reference standard. Where studies disagreed, the divergence was examined for differences in cohort composition, reference standard, phenotype ascertainment and outcome definition before any interpretive explanation was advanced. Formal risk-of-bias scores were not assigned, because the recognised instruments for prediction-model appraisal were not applied in the primary studies and the reporting in most of them is insufficient to support post hoc scoring.

3. The Analytical Substrate: Why the Fetal Genome Resists Conventional Computation

3.1 The Mixed-signal Problem in Maternal Plasma

Every noninvasive assessment of the fetal genome begins from an inconvenient fact. The material sampled is not fetal DNA but a mixture in which fetal-derived sequences form a minority component against a maternal background (Lo *et al*., 1997). The two populations are not separable by any physical property that admits clean partition, and inference must therefore proceed by exploiting subtle statistical differences between them. Sequencing of maternal plasma at sufficient depth allows reconstruction of the genome-wide fetal genetic and mutational profile, but only through inference from allele ratios and parental haplotype structure rather than through direct observation of a fetal genome (Lo *et al*., 2010). The first noninvasive whole-genome sequence of a human fetus was assembled in exactly this way, from deeply sequenced maternal plasma combined with parental genotype and haplotype information (Kitzman *et al*., 2012).

Three consequences follow for any computational method operating on this substrate. First, the proportion of fetal-derived material, conventionally the fetal fraction, is itself an estimated quantity rather than a measured one, and every downstream inference is conditional on that estimate. Methods based on fragment length distributions, on genomic location of fragments, and on Y-chromosome representation in male pregnancies give different answers, and combining several attributes yields greater accuracy than any single attribute, with the greatest gains where the fetal fraction is low and the risk of an incorrect conclusion is highest (Gazdarica *et al*., 2019). Second, the signal that distinguishes fetal from maternal molecules is carried in properties of the fragments themselves, including their lengths and their genomic distributions, which is precisely the type of high-dimensional, weakly structured information for which learned representations are well suited. Third, because the reference against which reads are aligned is a population construct, polymorphism and reference bias contribute systematic error that is not reducible by increasing depth alone (Xue *et al*., 2024).

3.2 Placental Origin, Mosaicism and the Identity of What is Being Measured

A second complication is biological rather than statistical. Circulating cell-free DNA of pregnancy derives predominantly from the placenta, and chorionic villus material obtained invasively shares that origin. Genotypic discordance between placenta and fetus is therefore possible, and this discordance, together with maternal copy number variation, vanishing twin and occult maternal malignancy, accounts for a recognised proportion of discrepant screening results, which is why professional guidance treats a positive cell-free DNA result as a screening finding requiring diagnostic confirmation rather than as a diagnosis (Dungan *et al*., 2023). The detection of maternal cancer as an incidental consequence of fetal aneuploidy screening is a well-described instance of the same phenomenon, in which an anomalous genomic signal originates from neither the fetus nor the normal maternal genome (Goldlust & Bianchi, 2025).

This matters computationally because it undermines the assumption on which supervised learning depends, namely that the label used in training corresponds to the state being inferred. A model trained to predict fetal karyotype from plasma features is in fact trained on placental genomic states labelled with fetal outcomes, and the residual discordance sets a ceiling on achievable accuracy that no amount of architectural sophistication can penetrate. The same argument applies with different force to invasively obtained samples, since chorionic villus sampling interrogates placental tissue whereas amniocentesis interrogates cells of fetal origin.

3.3 The Phenotype as the Weakest Link

The third and most consequential feature of the fetal setting concerns the phenotype rather than the genome. Interpretation of a sequence variant is not a property of the variant alone; it depends on the clinical picture against which the variant is assessed, a dependence formalised in the criteria that underpin contemporary variant classification (Richards *et al*., 2015). In postnatal rare-disease diagnostics, the clinical picture is obtained by examination and can be extended by further examination when the initial

genomic result is uninformative. In a fetus, the phenotype is inferred from ultrasound and magnetic resonance imaging, is limited to features that are anatomically visible at the gestational age examined, and excludes entire domains of information that dominate postnatal syndrome recognition, including neurodevelopment, behaviour, tone, growth trajectory and most biochemical parameters.

The instability of this phenotype is empirically documented rather than merely theoretical. In a consecutive cohort of prenatal exome cases, additional anomalies were detected between the second and third trimesters in 73.3% of pregnancies in which a causative variant had been identified and which reached the third trimester, with progressive hydropic features, arthrogryposis and brain anomalies accounting for most of the additions, and three variants of uncertain significance were reclassified as likely pathogenic on the basis of postnatal information (Mone *et al*., 2022). Reanalysis of unsolved prenatal exomes with postnatal and post-mortem features added diagnoses that the prenatal phenotype had not supported (Thauvin-Robinet *et al*., 2025). The phenotype supplied to a prioritisation algorithm at the moment of prenatal analysis is therefore not merely incomplete but systematically incomplete in a direction that changes with gestational age.

The complementary consequence is that improving the depth of prenatal phenotyping does not automatically improve algorithmic performance. A retrospective analysis of a multicentre prenatal exome study found that deep phenotyping of fetal brain abnormalities, defined as targeted ultrasound supplemented by magnetic resonance imaging, autopsy or family phenotype information, was not associated with an increased diagnostic yield of trio exome sequencing, and that neural tube defects were significantly associated with negative results (Drexler *et al*., 2023). The cohort was small and the analysis exploratory, so the finding should not be over-read, but it does show that the relationship between phenotypic richness and diagnostic success in the fetus is not the simple monotonic one that phenotype-driven algorithms presuppose.

Table 1 sets out the analytical layers involved in in utero fetal genomic assessment, the computational task that dominates each, and the feature of the fetal setting that most constrains performance at that layer. The table is intended to make explicit a distinction that is often blurred in the literature, namely that the layers differ not in the sophistication of the methods available but in the quality of the reference standard against which those methods can be evaluated. Where a layer has an objective and stable reference standard, such as a confirmed karyotype or an orthogonally validated genotype, algorithmic performance can be measured meaningfully. Where the reference standard is itself a curated and revisable judgement, as in variant classification and candidate-gene ranking, reported performance measures the concordance of an algorithm with current expert opinion rather than with biological truth.

Table 1. Analytical layers of in utero fetal genomic assessment, their dominant computational tasks and the principal constraints imposed by the prenatal setting

Analytical layer	Principal data type	Dominant computational task	Principal constraint in the fetal setting	Supporting references
Fetal fraction estimation	Fragment lengths, genomic coordinates, chromosome Y representation	Regression on mixture proportion	Estimate is unobservable directly; accuracy degrades where fetal fraction is lowest and stakes are highest	Gazdarica *et al*. (2019); Lo *et al*. (1997)
Chromosomal aneuploidy detection from plasma	Low-coverage whole-genome read counts and fragment features	Binary or multi-class classification against a threshold	Placental rather than fetal origin of the signal; reference and polymorphism bias	Dungan *et al*. (2023); Lee *et al*. (2022); Xue *et al*. (2024)
Noninvasive sequence-variant determination	Deep or ultra-deep plasma sequencing, parental haplotypes	Probabilistic genotype inference in a mixture	Maternally inherited alleles are intrinsically harder to resolve than paternally inherited alleles	Kitzman *et al*. (2012); Liscovitch-Brauer *et al*. (2024); Schwammenthal *et al*. (2025)
Variant detection from invasive samples	Short-read or long-read alignments	Sequence-to-genotype classification	Comparable to postnatal practice; principal constraint is turnaround time rather than accuracy	Poplin *et al*. (2018); Ren *et al*. (2024)
Variant-effect prediction	Protein sequence, structure, evolutionary conservation, splice context	Ranking or scoring of functional consequence	Training and calibration derive from postnatal and predominantly adult-ascertained variant sets	Cheng *et al*. (2023); Jaganathan *et al*. (2019)
Phenotype-driven gene prioritisation	Ontology-encoded clinical features plus variant call set	Semantic similarity and combined ranking	Fetal phenotype is imaging-derived, gestationally incomplete and unstable	Cooperstein *et al*. (2025); Gargano *et al*. (2024); Mone *et al*. (2022)
Candidate-diagnosis nomination	Integrated genome plus phenotype representation	Multi-source evidence integration	No prospective prenatal evaluation with decision-relevant endpoints	Clark *et al*. (2019); De La Vega *et al*. (2021)

Note. The table summarises constraints identified in the cited literature and does not represent an exhaustive account of every computational task performed within a prenatal genomic service

4. Machine Learning in Cell-Free DNA Analysis: From Counting Statistics to Learned Representations

4.1 Aneuploidy Detection and the Problem of Diminishing Headroom

Screening for the common autosomal trisomies using circulating cell-free DNA was built on counting statistics. Reads are mapped, counts per chromosome are normalised, and a standardised deviation is compared with a threshold. This approach performs extremely well for trisomy 21 and acceptably for trisomies 18 and 13, as established by cumulative meta-analysis (Gil *et al*., 2017), and its performance in a general-risk population supports its use as a primary screening test (Dungan *et al*., 2023). Large service-level cohorts confirm that the approach is robust in routine practice; in a programme applying genome-wide testing as a second-line test after combined first-trimester screening in 12,700 pregnancies, the overall false-positive rate was 0.3%, and fetal fraction was significantly associated with maternal body mass index and weight as well as with fetal sex (Baranova *et al*., 2022).

The headroom available to a learned classifier is therefore narrow for the principal target and wider only at the margins. Those margins are nonetheless where clinical difficulty concentrates: low fetal fraction, borderline standardised scores, rare autosomal aneuploidies, sex chromosome aneuploidies and subchromosomal imbalances. Two lines of work illustrate what can and cannot be claimed. A convolutional neural network trained on a fragment-distance representation rather than on fragment counts was developed on 2,215 samples and evaluated on 17,678 clinical samples, where the ensemble model achieved sensitivity of 99.07% and positive predictive value of 88.43% for the three common trisomies, compared with 98.15% sensitivity and positive predictive values of 40.77% and 36.81% for conventional standardised-score and normalised chromosomal value approaches (Lee *et al*., 2022). A separate approach addressed the reference-bias component by incorporating population variation through a pan-genome model with sequence-to-graph alignment, and reported superiority over the standard standardised-score test on two simulated datasets and 745 real cell-free DNA datasets, with the authors positioning the method as an adjunct for samples close to the decision threshold rather than as a replacement for it (Xue *et al*., 2024).

These results merit careful reading. The gain in positive predictive value reported for the fragment-distance classifier is large, and if it were reproduced independently it would represent a material reduction in unnecessary invasive procedures. The comparison, however, was internal, conducted within a single laboratory on a single sample stream, and the comparator was implemented by the same group. Positive predictive value depends on the prevalence of the target condition in the tested population and on the criteria used to adjudicate a positive, so a like-for-like comparison across laboratories cannot be inferred from these figures. External validation in an independent population with an independently adjudicated reference standard has not been reported, and calibration, as distinct from discrimination and threshold-based accuracy, was not presented. The pan-genome approach is more explicit about its own scope, reporting improvement relative to a defined statistical baseline and framing its output as supplementary evidence for borderline samples, which is an appropriately restrained claim given the evaluation performed.

4.2 Fetal Fraction as a Learned Quantity

Fetal fraction estimation illustrates the case where machine learning is unambiguously appropriate. The quantity of interest is a mixture proportion that cannot be observed directly, several partially independent sources of evidence bear on it, and the reference standard for male pregnancies is objective. Combining fragment-length and fragment-count information in a single predictor improves accuracy relative to either source alone, and weighting the training process towards samples with low fetal fraction improves accuracy precisely where the clinical risk of a wrong conclusion is greatest (Gazdarica *et al*., 2019). This is a regression problem with a defensible reference standard and a clear rationale for information fusion, and the reported gains are modest, specific and plausible.

The unresolved difficulty is that the objective reference standard is available only in male pregnancies, since it depends on chromosome Y representation. Models are therefore trained on a subset of pregnancies and applied to all of them, and the assumption that the relationship between fragment features and mixture proportion is invariant with respect to fetal sex is rarely tested explicitly. This is a form of distributional shift between the development and deployment populations of exactly the kind that degrades clinical prediction models in practice (Finlayson *et al*., 2021). The observation that fetal fraction varies systematically with maternal body mass index and weight, and with fetal sex, in a large service cohort (Baranova *et al*., 2022) indicates that the covariate structure of the deployment population is not uniform, which strengthens rather than weakens the case for explicit subgroup evaluation.

4.3 Fragmentomics and Representation Learning

The most active methodological frontier concerns features of cell-free DNA fragments beyond their number and location, including end motifs, size distributions and inferred nucleosome positioning. Comprehensive appraisal of machine learning applied to cell-free DNA diagnostics identifies fragmentomic representation as the principal source of new information available to learned models, and describes the algorithmic families that have been applied to it (Tsui *et al*., 2025). Development of a deep learning model that classifies disease status from fragment end motifs demonstrates that such features carry recoverable signal, although that work was conducted in oncology, where the biological question is tissue of origin and where labelled outcome data are comparatively abundant (Shen *et al*., 2024).

Transfer of these approaches to the prenatal setting is plausible but not established. The oncological application seeks to detect a tumour-derived minority component against a haematopoietic background, and the prenatal application seeks to resolve a placental minority component against a maternal background, so the mathematical structure of the problems is similar. The biological structure is not. Tumour-derived fragments carry aberrant methylation and chromatin signatures that reflect somatic disorganisation, whereas placental-derived fragments carry a normal tissue-specific epigenetic programme, and the magnitude of the distinguishing signal cannot be assumed to be comparable. Reviews addressing the genomic applications of noninvasive prenatal testing describe this broader research programme without claiming that fragmentomic classification currently outperforms established methods for the standard prenatal indications (Jin *et al*., 2024).

4.4 Noninvasive Determination of Single-gene Disorders

The clinically most consequential extension of noninvasive testing concerns monogenic conditions. Established noninvasive prenatal diagnosis for single-gene disorders rests on relative mutation dosage and relative haplotype dosage, in which the balance of maternal alleles in plasma is compared against expectation under alternative fetal genotypes, and these techniques have moved into clinical service for defined indications (Hanson *et al*., 2022; Hanson *et al*., 2023). The asymmetry between paternally and maternally inherited alleles is intrinsic to the method, since a paternally inherited allele absent from the maternal genome is directly detectable whereas a maternally inherited allele must be inferred from a shift in a ratio.

Learning-based methods have been applied to that asymmetry. A method combining fragment-level features, Bayesian inference and a random forest refinement step trained on millions of non-pathogenic variants was evaluated on 16 families with singleton pregnancies across a range of fetal fractions, reporting mean areas under the receiver operating characteristic curve of 0.996 for paternally inherited variants, 0.81 for maternally inherited variants, 0.86 for homozygous biallelic variants and 0.95 for compound heterozygous variants, with correct prediction of affected or unaffected status in all ten families carrying a pathogenic single-nucleotide variant (Liscovitch-Brauer *et al*., 2024). A subsequent deep learning framework applied to ultra-deep whole-genome sequencing of plasma integrated nucleotide, fragment, region, sample and familial levels of information, reported improved performance over existing approaches and detected three deleterious mutations, with inference possible from the seventh week of gestation (Schwammenthal *et al*., 2025). A separate self-learning neural network was implemented on deep cell-free DNA sequencing data from ten fetuses with monogenic disease and structural anomalies as a proof of concept (Qi *et al*., 2024).

The pattern across these studies is consistent and should be stated plainly. The reported discrimination for the difficult maternally inherited case, at an area under the curve of approximately 0.81, is far below what would be required for a standalone prenatal diagnostic decision, whereas the near-perfect performance for paternally inherited variants restates a problem that was already tractable. The evaluations involve tens of families at most. Case-level success in all families containing a pathogenic variant is encouraging but rests on very small numbers, and with ten such families the upper bound of the confidence interval around a perfect result remains wide. The authors of the deep learning framework acknowledge the residual risk of information leakage when rare variants are shared across training and test families, and identify model explainability as a barrier to clinical adoption. None of these studies reports external validation in a cohort recruited and analysed independently, and none reports calibration or decision-curve behaviour.

4.5 What the Cell-free DNA Evidence Supports

The evidence supports three conclusions with reasonable confidence. Learned models can extract information from fragment-level features that count-based statistics discard, and this information is most useful at the decision margins rather than for the principal screening targets. Fetal fraction estimation is genuinely improved by information fusion, with the largest gains in the samples that matter most. Noninvasive determination of monogenic conditions genome-wide remains at proof-of-concept stage for maternally inherited alleles, notwithstanding the confident framing of some reports.

Two conclusions cannot be supported. There is no published evidence that any learned cell-free DNA classifier reduces the rate of invasive procedures, the rate of inconclusive results or parental anxiety in a prospectively evaluated service, since no such evaluation has been reported. There is also no basis for treating reported positive predictive values as transferable across laboratories, because they are functions of the prevalence and adjudication practice of the cohort in which they were measured.

Table 2 summarises representative applications of machine learning to cell-free DNA-based fetal genomic analysis. Read as a whole, the table demonstrates a systematic pattern in which the strength of the analytical claim is inversely related to the size and independence of the evaluation. The two studies with the largest evaluation sets address the task with the least clinical headroom, namely common trisomy detection, whereas the studies addressing the tasks of greatest potential clinical value, namely genome-wide noninvasive determination of monogenic disease, rest on cohorts of ten to sixteen families. This inversion is the central methodological weakness of the field, and it is not remedied by the technical quality of the individual studies, which is generally high.

Table 2. Representative applications of machine learning to cell-free DNA-based analysis of the fetal genome

Application	Approach as reported	Scale of evaluation	Reported performance	Principal limitation for prenatal inference	Supporting references
Common trisomy detection	Convolutional neural network on fragment-distance representation	2,215 development samples; 17,678 clinical samples	Sensitivity 99.07%; positive predictive value 88.43% for the ensemble model	Single-centre internal comparison; positive predictive value is cohort-dependent; calibration not reported	Lee *et al*. (2022)
Aneuploidy detection with reduced reference bias	Pan-genome model with sequence-to-graph alignment feeding a standardised score	Two simulated datasets; 745 clinical datasets	Superior to the standard standardised-score test; framed as adjunct for borderline samples	No independent external cohort; benefit confined to samples near the threshold	Xue *et al*. (2024)
Fetal fraction estimation	Combination of fragment-length and fragment-count predictors with weighted training	2,454 euploid male-fetus samples	Improved accuracy over single-attribute predictors, particularly at low fetal fraction	Objective reference standard available only in male pregnancies	Gazdarica *et al*. (2019)
Genome-wide noninvasive determination of inherited variants	Fragment features with Bayesian inference and random forest refinement	16 families	Area under the curve 0.996 paternal, 0.81 maternal, 0.86 homozygous biallelic, 0.95 compound heterozygous	Discrimination for maternally inherited alleles inadequate for standalone diagnosis; very small cohort	Liscovitch-Brauer *et al*. (2024)
Deep learning fetal genotyping from ultra-deep plasma sequencing	Multi-level neural framework spanning nucleotide, fragment, region, sample and family	Small family cohort with ultra-deep whole-genome sequencing	Improved on prior methods; three deleterious mutations detected; feasible from the seventh week	Acknowledged leakage risk from shared rare variants; explainability identified as an adoption barrier	Schwammenthal *et al*. (2025)
Neural network detection of fetal single-nucleotide variants	Self-learning neural network on deep cell-free DNA sequencing	10 fetuses with monogenic disease	Reported as proof of concept	Very small cohort; brief report format limits methodological assessment	Qi *et al*. (2024)
Fragmentomic representation learning	Deep models on fragment end motifs and related features	Oncological cohorts	Recoverable disease signal from end-motif features	Developed outside the prenatal setting; magnitude of the analogous placental signal not established	Shen *et al*. (2024); Tsui *et al*. (2025)

Note. Performance figures are those reported by the cited authors and are not directly comparable across rows, because the cohorts, reference standards and target conditions differ

5. Deep Learning in the Diagnostic Prenatal Genome: Detection and Effect Prediction

5.1 Variant Detection is Largely a Solved Engineering Problem

Where fetal DNA is obtained invasively, the computational problem of identifying variants is substantially the same as in any other diagnostic context. Reformulating variant calling as an image-classification task over pileup representations, and training a deep neural network to perform it, produced a caller that generalised across sequencing platforms and genome builds without task-specific retuning (Poplin *et al*., 2018). Subsequent development has extended learned callers to long reads, to somatic and mosaic contexts, and to joint processing of multiple data types, and the resulting body of work has been reviewed comprehensively (Ren *et al*., 2024). Benchmarking against reference materials is mature, the reference standard is objective, and performance is high.

This layer therefore contributes little to the specifically prenatal difficulty, and it is important not to allow its success to be transferred rhetorically to layers where the evidence is far weaker. The prenatal constraint at the detection layer is turnaround time rather than accuracy. Service evaluations of rapid prenatal exome sequencing describe pathways engineered around days rather than weeks (Walton *et al*., 2025), and the same cohort analysis that documented evolving fetal phenotypes reported a reduction in mean turnaround time from 54.0 days to 14.2 days following the adoption of a defined national pathway with explicit inclusion criteria and a prenatally focused gene panel (Mone *et al*., 2022). Computational acceleration contributes to that reduction, but the larger contributors were organisational: defined eligibility, a curated panel and a structured multidisciplinary review.

5.2 Missense and Splicing Effect Prediction: Strong Tools, Uncertain Prenatal Calibration

The interpretive layer is where deep learning has made its most visible contribution and where prenatal application is most problematic. A deep neural network predicting splicing directly from primary sequence achieved a substantial advance over prior splice predictors and identified cryptic splice variants of likely clinical relevance (Jaganathan *et al*., 2019). A proteome-wide missense effect predictor built on protein structure prediction and population variation subsequently provided classifications across the majority of possible human missense substitutions (Cheng *et al*., 2023). Both are now embedded in routine diagnostic filtering.

Independent evaluation in defined disease areas has been more equivocal than the headline claims suggest. Benchmarking a proteome-wide missense predictor against curated clinical classifications in inherited retinal disease showed useful but imperfect concordance, with a substantial number of variants classified differently by the two sources (Lindquist *et al*., 2026). Evaluation in the diagnostic setting of neuromuscular disorders similarly found the predictor informative but not sufficient as a standalone criterion for classification (Krenn *et al*., 2026). For splicing, retraining on curated alternative splice sites materially altered predictions relative to the original model, indicating sensitivity to the composition of the training annotation rather than to the underlying biology alone (Strauch *et al*., 2022), and a systematic assessment of splicing variant prediction across task-specific deep models and genomic foundation models found that performance varies substantially by task formulation and evaluation design (Sun *et al*., 2026).

Three properties of these tools bear directly on prenatal use. First, they are trained and calibrated predominantly on variant sets ascertained in postnatal populations, in which conditions incompatible with survival to birth are systematically under-represented. Constraint metrics derived from large population reference datasets share this ascertainment structure, since they are computed from individuals who were born and recruited (Chen *et al*., 2024). The prior probability that a variant in a developmentally essential gene is deleterious is not the same in a fetal cohort as in an adult reference population, and no published recalibration of these predictors to prenatally ascertained variant sets exists. Second, classification frameworks assign in silico evidence a supporting rather than a determinative weight (Richards *et al*., 2015), and this weighting was derived without specific consideration of the prenatal setting, where the phenotypic evidence that would ordinarily counterbalance computational evidence is weakest. Third, the predictors express confidence on scales that are not probabilities of clinical pathogenicity, and their outputs are frequently reported to clinicians without the calibration information that would permit appropriate discounting.

5.3 The Prenatal Calibration Problem

The core difficulty can be stated compactly. Variant-effect predictors estimate the probability that a variant disrupts molecular function. Prenatal decision-making requires the probability that a fetus carrying that variant will have a particular clinical outcome, which is a different quantity, related to the first through penetrance, expressivity, modifier effects and gestational timing. The gap between these quantities is bridged in postnatal practice by observed phenotype. In the fetus, that bridge is largely absent, and its absence is greatest for the neurodevelopmental outcomes that most often drive parental decision-making.

The practical expression of this problem is the variant of uncertain significance. Practice varies substantially between laboratories in whether and how such variants are reported prenatally (Cornthwaite *et al*., 2022), and the question of whether reporting them serves patients has been debated on both empirical and ethical grounds, with arguments that selective reporting can support later reclassification and arguments that it imposes uninterpretable burdens at a moment of maximum vulnerability (de Koning *et al*., 2025; Gibbs *et al*., 2026). Any computational method that increases the number of variants surfaced for consideration without correspondingly increasing the strength of evidence attached to them will increase the volume of this problem. This is a foreseeable consequence of deploying high-sensitivity prioritisation in a setting with a weak phenotype, and it is rarely acknowledged in reports of algorithmic performance.

Table 3 gathers the recurrent methodological limitations identified across the studies reviewed in this and the preceding section. The pattern it displays is not one of poor science but of a mismatch between the evaluation designs that the field has adopted and the inferential questions that prenatal practice poses. Discrimination is reported almost universally, calibration almost never; internal validation is standard, external validation exceptional; and the endpoint is nearly always concordance with an existing reference standard rather than any change in a clinical process or outcome. Until these gaps are addressed, claims that computational methods improve prenatal genomic care will remain inferences from analytical performance rather than demonstrations.

Table 3. Recurrent methodological limitations in the reviewed literature on computational analysis of the fetal genome

Limitation	How it manifests in the reviewed studies	Consequence for prenatal interpretation	Supporting references
Absence of external validation	Development and evaluation within a single centre, sample stream or family cohort	Reported performance may reflect local sequencing, population and adjudication practice rather than transferable model behaviour	Lee *et al*. (2022); Liscovitch-Brauer *et al*. (2024); Xue *et al*. (2024)
Discrimination reported without calibration	Areas under the curve, sensitivity and positive predictive value presented; predicted-versus-observed agreement absent	Clinicians cannot judge how far a numerical score should shift a prior probability in an individual pregnancy	Collins *et al*. (2024); Schwammenthal *et al*. (2025)
Very small evaluation cohorts for high-stakes tasks	Ten to sixteen families used to evaluate genome-wide noninvasive determination of monogenic disease	Confidence intervals around apparently perfect case-level performance remain wide	Liscovitch-Brauer *et al*. (2024); Qi *et al*. (2024)
Reference standard is expert judgement rather than biological truth	Concordance measured against curated variant classifications that are themselves revisable	Performance measures agreement with current consensus, and inherits its errors and gaps	Cheng *et al*. (2023); Lindquist *et al*. (2026); Richards *et al*. (2015)
Postnatal ascertainment of training and reference data	Population constraint metrics and clinical variant sets derived from individuals born and recruited	Prior probabilities embedded in the tools do not match the fetal population to which they are applied	Chen *et al*. (2024); Krenn *et al*. (2026)
Label-target mismatch in plasma-based inference	Placental genomic state labelled with fetal outcome	Sets an achievable ceiling on accuracy independent of model quality	Dungan *et al*. (2023); Goldlust & Bianchi (2025)
Analytical rather than clinical endpoints	Outcomes are concordance metrics, not diagnostic yield, turnaround time, procedure rates or decisions	Clinical utility is asserted by inference rather than demonstrated	Collins *et al*. (2024); Van den Veyver *et al*. (2022)
Opacity of model reasoning	Model outputs presented as scores without accessible justification	Impedes multidisciplinary review and informed consent in an irreversible decision context	Coghlan *et al*. (2024); Rudin (2019); Schwammenthal *et al*. (2025)

Note. Limitations are those identifiable from the published reports and, where explicitly acknowledged by the original authors, are attributed as such in the text

6. Phenotype-Driven Prioritisation and the Incomplete Fetal Phenotype

6.1 How Phenotype-driven Algorithms Work and Where Their Evidence Comes From

The dominant paradigm for reducing a genome to a reportable shortlist combines variant-level evidence with a computable representation of the patient's clinical features. Clinical features are encoded as terms from an ontology, most commonly the Human Phenotype Ontology, which organises the phenotypic abnormalities of human disease in a form that supports semantic similarity computation and machine learning (Gargano *et al*., 2024). The algorithm scores each candidate gene by combining the deleteriousness and inheritance plausibility of the variants it carries with the semantic similarity between the patient's terms and the known phenotypic profile of the associated disease.

The performance of this paradigm is strongly parameter-dependent, and the dependence is larger than is generally appreciated. A systematic evaluation across 386 diagnosed probands from an undiagnosed diseases programme found that parameter optimisation raised the proportion of coding diagnostic variants ranked within the top ten candidates from 49.7% to 85.5% for genome sequencing data and from 67.3% to 88.2% for exome sequencing data, and raised the equivalent figure for non-coding variants from 15.0% to 40.0% (Cooperstein *et al*., 2025). Two implications follow. Default configurations of widely used prioritisation software miss a substantial fraction of diagnostic variants within the range a human reviewer would examine, and the gap between default and optimised performance is comparable in magnitude to the differences typically claimed between competing tools. It is also notable that even after optimisation, three-fifths of non-coding diagnostic variants fell outside the top ten, which bears directly on the value of genome as opposed to exome sequencing in a time-limited prenatal pathway.

Integrated candidate-nomination systems have reported stronger results. A clinical decision support tool benchmarked retrospectively on 119 probands, mostly infants in neonatal intensive care, ranked more than 90% of causal genes first or second and prioritised a median of three candidate genes per case, with performance maintained when trios were analysed as singletons and when phenotype descriptions were derived automatically from clinical notes rather than curated manually, and results were replicated in a further 60 cases from five academic centres (De La Vega *et al*., 2021). An earlier platform combining rapid genome sequencing with automated phenotype extraction from electronic health records reported clinical natural language processing precision of 80% and recall of 93%, automated retrospective diagnoses concurring with expert interpretation at 97% recall and 99%

precision in 95 children, and prospective correct diagnosis of three of seven critically ill infants with a mean time saving of over 22 hours (Clark *et al*., 2019).

These are substantial achievements, and they are also the strongest available evidence bearing on the prenatal question. That is precisely the problem. Every one of these evaluations was conducted in postnatal cohorts, and the phenotype that drove the ranking was extracted from clinical notes describing examined children. The mechanism by which the systems achieve their performance is the richness of the phenotype available to them.

6.2 The Fetal Phenotype does not Behave Like a Postnatal Phenotype

Direct evidence on the behaviour of ontology-driven gene selection in fetuses is sparse but pointed. In a single-centre series in which the whole exome was sequenced in 206 pregnancies with selected fetal structural anomalies and interpreted against approximately 5,000 morbid genes, retrospective application of alternative selection strategies showed that a curated prenatal panel detected 76 of 79 causative genes, or 96%, whereas Human Phenotype Ontology-driven gene selection identified only 56, or 71%; for incidental findings, the curated panel captured 8 of 19 and ontology-driven terms captured 1 (Ardiles-Ruesjas *et al*., 2026). Nearly one-third of primary findings would have been missed by the ontology-driven route. This is a retrospective single-centre analysis with the limitations that entails, but it addresses the exact question that the postnatal benchmarking literature cannot address, and its direction is unambiguous.

Why ontology-driven selection should underperform in fetuses is readily explained. The ontology encodes phenotypic abnormalities as observed and described in living individuals, and its annotation base is correspondingly weighted towards postnatally observable features. A fetus presents ventriculomegaly, short long bones or increased nuchal translucency, which are nonspecific terms shared by many disorders, and does not present the specific features that discriminate between them. Semantic similarity computed over nonspecific terms conveys little information, and the ranking then depends almost entirely on variant-level evidence, which is the component whose prenatal calibration is weakest.

Independent evidence from a different angle points the same way. A systematic review and comparative diagnostic accuracy study applied five national sets of phenotypic eligibility criteria to a historical cohort of 1,054 unselected structurally abnormal fetuses undergoing prenatal exome sequencing. The English National Health Service criteria achieved a pooled sensitivity of 49.8% and specificity of 80.7%, with an area under the summary receiver operating characteristic curve of 0.66; the best-performing alternative, from British Columbia, achieved sensitivity of 70.5%, specificity of 68.9% and an area under the curve of 0.73; and the likelihood of a monogenic diagnosis did not increase significantly as the number of criteria met increased (Reilly *et al*., 2025). If expert-designed phenotypic criteria discriminate this weakly, the assumption that automated semantic matching over the same phenotypic information will discriminate strongly is not tenable.

The same study reported that the diagnostic yield of the gene panel adopted in 2024 was significantly higher than that of the original 2020 panel, 95.6% compared with 87.4% (Reilly *et al*., 2025). This is worth emphasising because it identifies where improvement actually came from. Gains in prenatal diagnostic performance over this period were driven by better curation of prenatally relevant gene content, not by better algorithms operating on unchanged content.

6.3 When Richer Phenotyping does and does not Help

The relationship between phenotypic depth and diagnostic success is not uniform across anomaly types. The exploratory analysis showing no association between deep phenotyping and exome yield in fetal brain abnormalities concerned a specific anomaly category in a cohort of 76 trios, and neural tube defects were significantly associated with negative results (Drexler *et al*., 2023). By contrast, a cohort of 352 fetuses with anomalies of the corpus callosum achieved an overall diagnostic yield of 23%, rising to 46% in fetuses with callosal dysgenesis, with pathogenic or likely pathogenic variants in 49 different genes (Héron *et al*., 2025). The reconciliation is that anatomical specificity, rather than the number of modalities used to describe an anomaly, drives yield. Callosal dysgenesis is a more discriminating description than a corpus callosum anomaly in general, and adding a magnetic resonance scan to an ultrasound that has already established a nonspecific finding may add little that alters gene ranking.

This distinction has a direct algorithmic implication that the current literature does not address. Phenotype-driven ranking benefits from specificity, not volume, yet the ontology term sets available for fetal imaging findings are shallower in exactly the anatomical domains where specificity is achievable. Improving the granularity of prenatal phenotype representation is a data-curation problem before it is a modelling problem.

6.4 Generative Models and the Counselling Interface

Interest has extended to generative language models as an interface between genomic output and clinical communication. An evaluation of four general-purpose generative systems on 102 questions concerning four rare diseases found that the systems generally produced reliable information, with mean expert ratings ranging from 3.35 to 4.24 on a five-point scale and statistically significant differences between systems, but that occasional inaccuracies and ambiguous referencing could cause confusion and anxiety, and that expert guidance on appropriate use is necessary (Jeon *et al*., 2025). A study of genetic counselling graduate programme leadership found mixed attitudes to curricular integration of these technologies, with respondents judging them least relevant to interpersonal, psychosocial and counselling competencies, and identifying an absence of resources, training and guidelines as the principal barrier (Feuer *et al*., 2025).

These findings are modest in scope, and neither addresses prenatal counselling specifically. They nonetheless indicate the shape of the risk. The task in prenatal genetic counselling is not primarily the transmission of accurate factual content; it is the calibrated communication of uncertainty in a setting where the recipient must make an irreversible decision under time pressure. A system that produces fluent, confident and mostly accurate prose is well suited to the first task and poorly suited to the second, and the failure mode is not a visible error but an unwarranted impression of certainty.

Table 4 sets out the principal points of disagreement in the literature about the value of computational prioritisation in prenatal practice, together with the most plausible explanation for each divergence. The pattern that emerges is instructive: almost every disagreement dissolves once the cohort composition and the source of the phenotype are specified. Positions that appear to conflict are usually answers to different questions asked of different populations, and the field would be considerably clarified if reports stated explicitly whether the phenotype driving an evaluation was examined, imaged or extracted from a record.

Table 4. Principal disagreements about computational prioritisation in prenatal genomic practice and their most plausible explanations

Point of disagreement	Position supported by one body of evidence	Position supported by another body of evidence	Most plausible explanation for divergence	Supporting references
Value of ontology-driven gene selection	Semantic similarity over ontology terms substantially improves candidate ranking	Ontology-driven selection in fetuses misses approximately 29% of causative genes retained by a curated prenatal panel	Postnatal evaluations use examined phenotypes; fetal application uses nonspecific imaging-derived terms	Ardiles-Ruesjas *et al*. (2026); Cooperstein *et al*. (2025); Gargano *et al*. (2024)
Effect of deeper phenotyping on yield	Automated deep phenotyping from records improves ranking and shortens time to diagnosis	Deep phenotyping of fetal brain abnormalities was not associated with higher exome yield	Anatomical specificity rather than modality count drives yield; brain findings are often nonspecific prenatally	Clark *et al*. (2019); Drexler *et al*. (2023); Héron *et al*. (2025)
Adequacy of in silico variant evidence	Proteome-wide and splice predictors approach experimental discrimination on benchmark sets	Independent disease-area benchmarking shows useful but insufficient concordance with curated classifications	Benchmarks reward agreement with existing classifications; clinical use requires calibrated probability in an unrepresented population	Cheng *et al*. (2023); Krenn *et al*. (2026); Lindquist *et al*. (2026)
Source of recent gains in prenatal yield	Improved computational prioritisation	Improved curation of prenatally relevant gene content and pathway design	Direct comparison of successive gene panels within one service shows the panel change accounted for the improvement	Mone *et al*. (2022); Reilly *et al*. (2025)
Reporting of variants of uncertain significance	Reporting supports later reclassification and parental information needs	Reporting imposes uninterpretable burden and drives inconsistent practice between laboratories	Divergence reflects differing weight given to parental autonomy against harm from uncertainty, not differing evidence	Cornthwaite *et al*. (2022); de Koning *et al*. (2025); Gibbs *et al*. (2026)
Readiness of noninvasive monogenic testing	Deep learning brings genome-wide noninvasive testing close to clinical feasibility	Discrimination for maternally inherited alleles remains inadequate for standalone diagnosis	The optimistic framing generalises from paternally inherited cases, which were already tractable	Liscovitch-Brauer *et al*. (2024); Schwammenthal *et al*. (2025)

Note. Positions are attributed to bodies of evidence rather than to individual authors, and the explanations offered are interpretive rather than established

7. From Analytical Validity to Clinical Utility

7.1 What Prenatal Sequencing Achieves, and against what Baseline

Any claim that computational methods improve prenatal genomic care must be assessed against the performance of the pathway as currently constituted. That performance is well characterised. In fetuses with structural anomalies and normal karyotype and microarray, exome sequencing identifies a diagnostic variant in approximately one in ten cases in unselected prospective cohorts, at 8.5% of 610 fetuses in one and 10% of 234 trios in the other, with yield varying markedly by phenotype: 15.4% in multisystem anomalies, 15.4% in skeletal anomalies and 3.2% in isolated increased nuchal translucency (Lord *et al*., 2019; Petrovski *et al*., 2019). Selection therefore matters more than analysis. In fetuses that are structurally normal, the pooled incremental yield of exome sequencing over standard testing is 1.6%, of which variants associated with severe disease account for 0.5% (Sotiriadis *et al*., 2025).

Genome sequencing adds further yield, with a pooled incremental yield over quantitative fluorescence polymerase chain reaction and microarray of 16% in prenatal cohorts, rising to 32% where sequencing was performed in line with national prenatal exome criteria (Shreeve *et al*., 2024). Prospective parallel comparison in 96 parent-fetus trios found that copy-number variant sequencing plus trio exome sequencing achieved a combined diagnostic rate of 31.2%, and that genome sequencing recovered all of those diagnoses plus three additional cases, raising the rate to 34.4%, while detecting incidental findings in 10.5% of families including structural variants identified only by genome sequencing (Gao *et al*., 2026). The incremental gain from the most comprehensive available assay is therefore real but modest, and it is accompanied by an increase in findings whose management is not settled.

7.2 Yield is not Utility

The distinction that the literature most often elides is between finding a variant and changing what happens. Direct evidence on this point exists and is striking. In the corpus callosum cohort, 73% of couples receiving a genetic diagnosis other than a variant in a gene associated with favourable outcome chose termination of pregnancy, compared with 17% of couples without a genetic diagnosis, and all couples in whom a variant was identified in a gene associated with a favourable neurodevelopmental outcome continued the pregnancy to term (Héron *et al*., 2025). In a consecutive service cohort, identification of a causative pathogenic variant had a significant clinical impact in 78.3% of cases, most frequently affecting decisions about the course of the pregnancy and neonatal management (Mone *et al*., 2022). Qualitative work with participants who underwent prenatal genome sequencing in the absence of fetal anomalies found that almost all those with positive results experienced downstream consequences including termination, cascade testing, specialist referral and changes to pregnancy and neonatal management, and that negative results were valued for reassurance (Kelly *et al*., 2026).

These findings establish that prenatal genomic results change decisions. They also establish the standard against which any computational contribution must be judged. An algorithm that improves the ranking of a causative gene from third to first has not necessarily changed any decision, because a human reviewer examining a shortlist of ten would have found it either way. An algorithm that surfaces a variant of uncertain significance that would otherwise not have been reported may change a decision, and not necessarily for the better. The direction of the utility effect is not determined by the direction of the analytical effect.

7.3 Time, Cost and the Shape of the Service

Where computational methods have a clearer claim is in the resource dimension. Turnaround time is the binding constraint in prenatal genomics, and the reduction of mean turnaround from 54.0 days to 14.2 days achieved in one service followed the adoption of defined inclusion criteria, a prenatally focused gene panel and structured multidisciplinary review rather than any change in algorithm (Mone *et al*., 2022). The automated postnatal platform that reported a mean time saving of over 22 hours in critically ill infants demonstrates that automation can compress interpretation time when the phenotype is machine-extractable (Clark *et al*., 2019), but no equivalent prenatal demonstration exists.

Cost data are similarly instructive. In the English national service, prenatal exome sequencing cost £2,331 per referred case and £3,592 per case that proceeded to testing, with an incremental cost per additional diagnosis of £11,326 at a diagnostic yield of 35.3% in the referred population, and an annual service cost of £1.8 million; incremental costs to families were negligible (Smith *et al*., 2025). The dominant cost component in such a service is skilled interpretive labour combined with the clinical infrastructure surrounding it, which is where automation could in principle exert leverage. Implementation evaluation of the same service identified seven distinct local delivery models and found that staff time and capacity, communication and collaboration, and organisational infrastructure were the principal determinants of successful implementation (Walton *et al*., 2025). A computational intervention that reduced interpretive workload without reducing accuracy would address a documented bottleneck. That such an intervention would do so has not been tested prospectively in any prenatal service.

7.4 Where Algorithmic Contribution is Most Plausible

Three points in the pathway offer the most defensible case for computational contribution, and it is worth distinguishing them from the applications that attract most attention. The first is triage of borderline cell-free DNA results, where a second, methodologically distinct classifier could reduce the number of pregnancies proceeding to invasive testing on the basis of an equivocal screening result, and where the reference standard for evaluation is objective (Xue *et al*., 2024). The second is automated reanalysis, since diagnoses accrue when unsolved cases are revisited against updated knowledge, and reanalysis of unsolved prenatal exomes with additional phenotypic information yields additional diagnoses (Thauvin-Robinet *et al*., 2025); reanalysis is a scheduled, low-urgency task in which algorithmic error carries less immediate risk and in which the workload argument is strongest. The third is quality assurance rather than primary interpretation, in which an automated system flags discrepancies between a reported classification and the current state of curated evidence, functioning as a check on human interpretation rather than a substitute for it. None of these three has been evaluated prospectively in a prenatal setting, and all three are more tractable than the applications currently receiving most attention.

8. Equity, Governance and the Ethics of Algorithmic Fetal Genomics

8.1 Representational Bias is Inherited, not Introduced

The reference resources on which every variant-interpretation algorithm depends do not represent human populations evenly. Population allele frequency and constraint resources are assembled from cohorts whose ancestral composition reflects the history of

genomic research participation, and this composition determines which variants appear rare, which appear common and which appear in a gene judged intolerant of variation (Chen *et al*., 2024). Curated clinical variant classification aggregates submissions from laboratories serving particular populations (Rehm *et al*., 2015; Richards *et al*., 2015). An algorithm trained or calibrated on these resources reproduces their structure. The consequence is predictable and directional: individuals from under-represented ancestral backgrounds receive more results of uncertain significance and fewer confident classifications, and any automated system that amplifies the throughput of interpretation amplifies this disparity in proportion.

This is not a hypothetical concern about algorithmic fairness. Analysis of a widely deployed health algorithm demonstrated that a model can encode substantial racial bias without any protected characteristic being used as an input, because the chosen label carried the bias (Obermeyer *et al*., 2019). The parallel to genomic interpretation is close. The label used to train and evaluate variant-effect predictors is the curated classification, and the curated classification is itself a product of who has been sequenced and whose variants have been submitted and reviewed. Reported model accuracy will look best in the populations best represented in the reference resources, and stratified reporting of performance by ancestry is almost entirely absent from the studies reviewed here.

In the prenatal setting the consequence is more than an inequitable distribution of uncertainty. An uncertain result at 20 weeks of gestation is not an invitation to further investigation over subsequent months; it is an input into a decision that must be made within weeks. Practice already varies substantially between laboratories in the prenatal reporting of uncertain variants (Cornthwaite *et al*., 2022), and unequal generation of such variants across ancestral groups therefore translates into unequal quality of prenatal counselling.

8.2 Opacity, Autonomy and the Structure of Consent

Informed consent for prenatal genomic testing requires that a pregnant person understand what the test can and cannot determine. Where a shortlist has been generated by a model whose reasoning is not accessible, the basis on which certain possibilities were excluded cannot be conveyed. The argument that high-stakes decisions should rely on interpretable models rather than on post hoc explanations of opaque ones (Rudin, 2019) applies with unusual force here, because the decision is irreversible, the timescale is short and the person bearing the consequences is not the person whose genome was analysed. Developers of prenatal genotyping models have themselves identified explainability as a barrier to clinical adoption (Schwammenthal *et al*., 2025).

Ethical analysis of artificial intelligence in prenatal and paediatric genomic medicine has framed the central issues as the effect of automation on reproductive autonomy, the opacity of automated recommendations, and the implications of expanded prenatal prediction for people living with the conditions being predicted (Coghlan *et al*., 2024). The last of these deserves particular emphasis because it is systematically under-discussed in the technical literature. An algorithm optimised to maximise the detection of pathogenic variants encodes an implicit judgement that detection is beneficial, and that judgement is contested by many whose conditions are the targets of detection. Professional guidance on prenatal sequencing has consistently located decision-making in a multidisciplinary process with counselling before and after testing (Monaghan *et al*., 2020; Van den Veyver *et al*., 2022), and automation that shortens the interpretive step without preserving the deliberative structure would erode the safeguard while leaving the appearance of it intact.

The adjacent debate on polygenic embryo screening illustrates where an unchecked predictive logic leads. Analysis of the expected gains from selecting embryos on polygenic scores concluded that predicted benefit is small, highly uncertain and dependent on assumptions that do not hold across ancestries, and that the practice raises serious ethical concerns (Turley *et al*., 2021). The technical structure of that argument transfers directly to any proposal to extend fetal genomic prediction from monogenic conditions to complex traits.

8.3 Privacy in a Tri-personal Context

Prenatal genomic analysis generates data about at least three people: the pregnant person, the fetus and, in trio analysis, a second genetic parent. Only one of these can consent contemporaneously, one will acquire rights over the data on reaching adulthood, and the third may have interests that diverge from the first. Genomic data are not reliably de-identifiable, as demonstrated by the recovery of surnames and identities from research genomes using genealogical resources (Gymrek *et al*., 2013), and the aggregation of large training corpora for model development compounds the exposure. The problem is structural rather than technical: a fetal genome sequenced in 2026 will describe an adult in 2044, and consent given by a parent in 2026 cannot anticipate the analytical capabilities or the data environment of that later period.

8.4 Regulation, Reporting and the Standards Gap

Regulatory frameworks for health artificial intelligence are being constructed around software as a medical device, with continuing uncertainty about the oversight of adaptive systems, the allocation of liability and the extent of pre-market evidence required (Mello & Cohen, 2025). Prenatal genomic interpretation sits awkwardly within these frameworks, since many of the components in use are open-source research tools embedded in laboratory pipelines rather than commercial devices, and are therefore governed, if at all, through laboratory accreditation rather than device regulation.

The reporting gap is more immediately actionable than the regulatory one. Contemporary guidance for reporting clinical prediction models developed with regression or machine learning specifies the items required to judge transferability, including the provenance and representativeness of the development data, the handling of missing data, the reporting of calibration alongside discrimination,

the specification of the intended use population and the evaluation of performance in relevant subgroups (Collins *et al*., 2024). Assessed against these items, the prenatal literature reviewed here falls short consistently: calibration is rarely reported, intended-use populations are seldom specified, subgroup performance is almost never presented, and external validation is exceptional. Adoption of these standards would not by itself demonstrate clinical utility, but it would allow readers to judge whether reported performance is likely to survive contact with a different population, which at present they cannot.

Table 5 translates the preceding analysis into prioritised research needs. The priorities are ordered by the ratio of expected benefit to methodological difficulty rather than by scientific interest, and several of the highest-ranked items require curation and study design rather than algorithmic innovation. This ordering reflects a judgement supported throughout this review, namely that the binding constraints on computational analysis of the fetal genome are the quality of the reference data and the design of the evaluations, not the capability of the available models.

Table 5. Prioritised research needs in computational analysis of the fetal genome

Priority	Unresolved problem	Appropriate design or methodological improvement	Principal outcome measures	Supporting references
Prenatal recalibration of variant-effect evidence	Predictors and constraint metrics are calibrated on postnatally ascertained populations	Assemble prenatally ascertained variant sets with outcome linkage; recalibrate and report predicted-versus-observed agreement	Calibration slope and intercept; reclassification rates in fetal cohorts	Chen *et al*. (2024); Cheng *et al*. (2023); Krenn *et al*. (2026)
Gestation-aware phenotype representation	Ontology terms for fetal imaging findings are shallow and lack temporal structure	Extend ontology annotation to gestational-age-specific findings; encode phenotype trajectories rather than single time points	Depth and specificity of applicable terms; effect on candidate-gene ranking	Ardiles-Ruesjas *et al*. (2026); Gargano *et al*. (2024); Mone *et al*. (2022)
Prospective evaluation with decision-relevant endpoints	Almost all evaluation uses concordance metrics rather than clinical endpoints	Prospective comparative studies embedded in prenatal services, with pre-specified endpoints and blinded adjudication	Turnaround time; invasive procedure rate; uncertain-result rate; decisional conflict	Collins *et al*. (2024); Smith *et al*. (2025); Walton *et al*. (2025)
External and multi-ancestry validation	Development and evaluation are typically confined to one centre and one population	Multi-centre validation with pre-registered protocols and mandatory ancestry-stratified reporting	Discrimination and calibration by ancestry group; rate of uncertain classifications by group	Chen *et al*. (2024); Obermeyer *et al*. (2019); Rehm *et al*. (2015)
Resolution of maternally inherited alleles in plasma	Discrimination for maternally inherited variants remains around 0.81 in the reported literature	Larger multi-centre family cohorts with orthogonal confirmation; explicit handling of information leakage across families	Area under the curve and calibration for maternally inherited alleles stratified by fetal fraction	Liscovitch-Brauer *et al*. (2024); Schwammenthal *et al*. (2025)
Automated reanalysis of unsolved prenatal cases	Reanalysis yields additional diagnoses but is labour-intensive and inconsistently performed	Trials of scheduled automated reanalysis with defined triggers and human confirmation	Additional diagnoses per unit of analyst time; time from knowledge update to case re-review	Cooperstein *et al*. (2025); Thauvin-Robinet *et al*. (2025)
Communication of algorithmic uncertainty	Model outputs reach counselling without calibration information	Development and testing of uncertainty presentations designed for prenatal counselling, evaluated with parents	Comprehension; decisional conflict; concordance between stated and understood certainty	Coghlan *et al*. (2024); Jeon *et al*. (2025); Kelly *et al*. (2026)

Note. Priorities are ordered by the ratio of expected benefit to methodological difficulty as judged in this synthesis, and the ordering is offered as a defensible proposal rather than as a consensus position

9. Future Directions

9.1 Building a Prenatally Ascertained Evidence Base for Variant Interpretation

The most consequential deficiency identified in this review is that the interpretive tools applied to fetal genomes carry calibration derived from populations that exclude much of the fetal phenotypic spectrum. Population constraint and allele frequency resources are assembled from individuals who were born and recruited (Chen *et al*., 2024), and clinical variant classifications accumulate from postnatally ascertained cases (Rehm *et al*., 2015). Conditions that are lethal before or shortly after birth, and conditions whose prenatal presentation differs from their postnatal presentation, are correspondingly under-represented.

The remedy is a prenatally ascertained resource linking variants to fetal phenotype and to pregnancy and childhood outcome, assembled prospectively across multiple centres and multiple ancestral populations. Existing prenatal sequencing services already generate the constituent data, and national services with defined pathways provide the necessary infrastructure (Walton *et al*., 2025). What is missing is systematic outcome linkage and a governance framework permitting aggregation. The immediate deliverable would be recalibration of existing predictors against prenatally ascertained variants, reported as calibration slope and intercept rather than as discrimination alone, with explicit quantification of how far current scores over- or under-state pathogenicity in a fetal population. The relevant scale is thousands of pregnancies with linked outcomes, which is achievable within five years through coordination of existing services rather than through new recruitment.

9.2 Representing the Fetal Phenotype as a Trajectory

Present practice encodes the fetal phenotype as a static set of ontology terms recorded at one moment. The empirical observation that anomalies accumulate through gestation in a large majority of pregnancies with a genetic diagnosis (Mone *et al*., 2022) means that this representation discards information about the direction and rate of phenotypic change, which is precisely the information a clinician uses when forming a differential diagnosis.

A gestation-aware representation would encode each finding with its gestational age at detection, its trajectory across serial examinations and its expected age of first detectability given the underlying condition. Ontology development would need to extend the annotation of prenatally observable features and to add temporal qualifiers, building on established infrastructure for phenotype representation and exchange (Gargano *et al*., 2024). The reason to prioritise this work is that it addresses the demonstrated failure mode of ontology-driven selection in fetuses, in which nearly one-third of causative genes were missed relative to a curated prenatal panel (Ardiles-Ruesjas *et al*., 2026). Curation of prenatal gene content has already been shown to deliver measurable improvements in yield within a single service (Reilly *et al*., 2025), which suggests that further curation effort will be productive. The appropriate evaluation is a retrospective application of the enriched representation to cohorts with known diagnoses, measuring the rank of the causative gene, followed by prospective assessment once the retrospective gain is established.

9.3 Prospective Evaluation against Endpoints that Matter

No prospective study has tested whether an artificial intelligence component changes any outcome in a prenatal genomic service. This is the central evidential gap, and it is not remediable by further analytical benchmarking.

The appropriate design is a prospective comparative study embedded in an operating service, in which cases are allocated to computationally assisted or standard interpretation with blinded adjudication of the final report. Endpoints should be pre-specified and clinically meaningful: turnaround time from sample receipt to report, diagnostic yield, the rate of variants of uncertain significance reported, the rate of invasive procedures following an equivocal screening result, analyst time per case, and decisional conflict among parents. Reporting should follow contemporary standards for prediction-model studies, including calibration, intended-use population and subgroup performance (Collins *et al*., 2024). The economic dimension is tractable within the same design, since the cost structure of prenatal sequencing services has been characterised in sufficient detail to support incremental analysis (Smith *et al*., 2025), and the marginal value of an intervention that reduces interpretive labour can therefore be estimated directly.

Two aspects of the design require particular care. Blinding is difficult because interpreters can usually tell whether a shortlist was machine-generated, so adjudication of the final report by an independent panel is preferable to blinding of the interpreter. Sample size is governed by turnaround time and analyst workload rather than by diagnostic yield, since yield differences would require impracticably large cohorts whereas process endpoints are detectable in hundreds of cases.

9.4 Reducing Inequity as a Design Objective Rather than a Caveat

Ancestry-stratified reporting of model performance should become a condition of publication rather than an optional addition. The mechanism by which a model can encode substantial group-level bias without using any protected characteristic as an input is well established (Obermeyer *et al*., 2019), and in genomic interpretation the mechanism operates through the composition of the reference resources themselves (Chen *et al*., 2024). Two concrete steps are available now. Studies should report the rate of uncertain classifications by ancestry group alongside overall accuracy, since that rate is the operative harm in prenatal practice. Services should record and publish the ancestral composition of the populations in which their pipelines were validated, so that a clinician can judge whether a given result is drawn from a well-represented or a poorly represented group. Neither step requires new methodology.

9.5 Longer-term Priorities

Three directions are important but less immediately tractable. The first is noninvasive determination of monogenic conditions genome-wide, where the resolution of maternally inherited alleles remains the limiting problem and where reported discrimination is well short of diagnostic adequacy (Liscovitch-Brauer *et al*., 2024; Schwammenthal *et al*., 2025); progress requires larger multi-centre family cohorts with orthogonal confirmation and explicit control of information leakage between training and evaluation families, on a timescale of years rather than months. The second is the integration of imaging and genomic evidence within a single inferential framework, so that quantitative image features contribute to gene ranking directly rather than through the lossy intermediary of ontology terms; the imaging component is technically mature (Arnaout *et al*., 2021; Horgan *et al*., 2023; Ramirez

Zegarra & Ghi, 2023), and the obstacle is the absence of paired image and genome datasets of sufficient size, which is a data-governance problem before it is a modelling problem. The third is the systematic evaluation of language models at the counselling interface, where current evidence indicates generally reliable factual content accompanied by occasional inaccuracy and ambiguous referencing (Jeon *et al*., 2025), and where the specific question for prenatal practice is not factual accuracy but whether such systems convey uncertainty faithfully; the workforce dimension of this question is already being examined within genetic counselling education (Feuer *et al*., 2025).

A final priority concerns what should not be pursued. Extension of fetal genomic prediction from monogenic conditions to complex traits and behavioural phenotypes lacks a defensible evidential basis, as the analysis of expected gains from polygenic embryo selection has shown (Turley *et al*., 2021), and computational sophistication does not supply the missing predictive validity. Research effort directed at that extension is unlikely to yield clinical benefit and carries substantial risk of harm.

10. Conclusions

Computational learning methods now operate at every layer of fetal genomic analysis, and the evidence supporting them varies more between layers than between techniques. Where the task is signal detection against an objective reference standard, the evidence is credible. Learned models extract usable information from fragment-level properties of cell-free DNA that count-based statistics discard, estimation of the fetal fraction is genuinely improved by combining partially independent sources of evidence, and deep neural variant calling from sequence alignments is a mature technology whose performance is not in serious dispute. These conclusions rest on internally consistent studies with defensible reference standards, although almost none has been externally validated in an independent prenatal population.

Where the task is inference about clinical meaning, the evidence is substantially weaker, and the weakness is structural rather than incidental. Variant-effect predictors and phenotype-driven prioritisation systems achieve their reported performance in postnatal cohorts, where the phenotype has been examined and can be re-examined. The fetal phenotype is imaging-derived, restricted to anatomically visible features, systematically incomplete in a direction that changes with gestational age, and often too nonspecific to discriminate between candidate genes. Direct evidence from prenatal cohorts shows that ontology-driven gene selection misses a substantial minority of causative genes that curated prenatal panels retain, and that expert-designed phenotypic eligibility criteria discriminate only moderately. The most defensible reading is that the recent improvement in prenatal diagnostic performance came from better curation of prenatally relevant gene content and better pathway design, not from better algorithms.

A further conclusion follows about the relationship between analytical performance and clinical benefit. Prenatal genomic results demonstrably change decisions, including decisions about the continuation of pregnancy, and the magnitude of that effect is large. It does not follow that improvements in analytical metrics change decisions, and in one important respect the relationship may be inverse, since methods that increase the number of variants surfaced without correspondingly increasing the strength of evidence attached to them will increase the volume of uncertain results in a setting where uncertainty is least tolerable. No prospective study has tested whether an artificial intelligence component alters turnaround time, diagnostic yield, the rate of uncertain results, the rate of invasive procedures or parental decision-making in a prenatal service. Claims of clinical utility in this field are therefore inferences from analytical validity rather than demonstrations, and they should be described as such.

The equity dimension is not a secondary consideration. The reference resources underpinning every interpretive algorithm represent human populations unevenly, and automation amplifies whatever structure those resources contain. In prenatal practice the resulting harm takes a specific and severe form, because an uncertain result delivered mid-gestation is not an invitation to further investigation but an input into a decision that must be made within weeks. Ancestry-stratified reporting of performance is almost entirely absent from the literature reviewed, and its absence should be treated as a reporting failure rather than as an acceptable limitation.

Taken together, the evidence supports a position that is neither dismissive nor promotional. Computational methods are already indispensable to prenatal genomic analysis at the detection layer and are contributing usefully at the margins of cell-free DNA screening. They are not yet dependable at the interpretive layer in the specific conditions the fetus presents, and the obstacles are the quality of the reference data and the design of the evaluations rather than the capability of the models. The most valuable near-term work is therefore curatorial and evaluative: prenatally ascertained variant evidence, gestation-aware phenotype representation, prospective studies with decision-relevant endpoints, and reporting sufficient to allow a reader to judge whether a result will transfer. Progress on those fronts would convert a set of plausible expectations into a body of evidence adequate to the gravity of the decisions it would inform.

11. Limitations

This review is a critical narrative synthesis and carries the limitations intrinsic to that design. Study selection was purposive rather than exhaustive, and although the search was structured, documented and applied across multiple sources, the possibility remains that relevant work was not retrieved or was retrieved and judged less central than another reader would judge it. Interpretive bias cannot be excluded, since the weighting of evidence across heterogeneous study types reflects the reasoning set out in the methods rather than any external adjudication, and a different reviewer applying the same criteria might reach different emphases. No second reviewer independently screened records or appraised included studies.

Access to subscription-indexed databases was not available, and searching was therefore confined to freely accessible biomedical and multidisciplinary sources supplemented by citation tracking. Work indexed only in subscription databases, or published in

journals not covered by the sources searched, may have been missed. The restriction to English-language publications is a further constraint, and it is likely to have reduced coverage of prenatal genomic practice in settings where the relevant literature is published in other languages, which compounds the geographical unevenness discussed in the synthesis.

The evidence base itself imposes limitations that no review design can overcome. Study designs, target conditions, reference standards and outcome definitions differ so substantially across the included material that quantitative synthesis was not possible, and comparisons between reported performance figures are therefore qualitative. Several of the most directly relevant studies rest on very small cohorts, and confidence intervals around their point estimates are correspondingly wide even where those estimates appear favourable. Formal risk-of-bias scoring was not applied, because the recognised instruments for appraising prediction-model studies were not used in the primary reports and the available reporting is insufficient to support reliable post hoc scoring.

Geographical representation among the included studies is uneven, with prenatal sequencing cohorts concentrated in a small number of high-income health systems, and this concentration limits the generalisability of the conclusions to settings with different testing pathways, different population structures and different resource constraints. Long-term outcome data are scarce throughout, since most prenatal genomic cohorts report diagnostic results rather than childhood outcomes, which restricts what can be said about the clinical consequences of any interpretive strategy. The field is also developing quickly, and some of the technical claims examined here will be superseded; the conclusions about evidential structure are more likely to persist than the conclusions about the performance of any particular tool. Finally, the absence of prospective evaluations with clinical endpoints means that several conclusions in this review are statements about what has not been demonstrated rather than about what has been refuted, and the distinction between absence of evidence and evidence of absence has been maintained throughout.

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

Ardiles-Ruesjas, V., Rodriguez-Revenga, L., Pauta, M., Nadal, A., Arca, G., Gort, L., Sinisterra-Díaz, S. E., & Borrell, A. (2026). Gene list selection matters: Missed diagnoses in prenatal exome sequencing—PanelApp R21 and HPO-driven versus OMIM-based gene lists. *Prenatal Diagnosis, 46*(5–6), 636–642. https://doi.org/10.1002/pd.70066

Arnaout, R., Curran, L., Zhao, Y., Levine, J. C., Chinn, E., & Moon-Grady, A. J. (2021). An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nature Medicine, 27*(5), 882–891. https://doi.org/10.1038/s41591-021-01342-5

Baranova, E. E., Sagaydak, O. V., Galaktionova, A. M., Kuznetsova, E. S., Kaplanova, M. T., Makarova, M. V., Belenikin, M. S., Olenev, A. S., & Songolova, E. N. (2022). Whole genome non-invasive prenatal testing in prenatal screening algorithm: Clinical experience from 12,700 pregnancies. *BMC Pregnancy and Childbirth, 22*(1), 633. https://doi.org/10.1186/s12884-022-04966-8

Chen, S., Francioli, L. C., Goodrich, J. K., Collins, R. L., Kanai, M., Wang, Q., Alföldi, J., Watts, N. A., Vittal, C., Gauthier, L. D., Poterba, T., Wilson, M. W., Tarasova, Y., Phu, W., Grant, R., Yohannes, M. T., Koenig, Z., Farjoun, Y., Banks, E., … Karczewski, K. J. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. *Nature, 625*(7993), 92–100. https://doi.org/10.1038/s41586-023-06045-0

Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., & Avsec, Ž. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science, 381*(6664), eadg7492. https://doi.org/10.1126/science.adg7492

Clark, M. M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., Ellsworth, K., Camp, B., Kint, C. I., Yacoubian, C., Farnaes, L., Bainbridge, M. N., Beebe, C., Braun, J. J. A., Bray, M., Carroll, J., Cakici, J. A., Caylor, S. A., Clarke, C., … Kingsmore, S. F. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Science Translational Medicine, 11*(489), eaat6177. https://doi.org/10.1126/scitranslmed.aat6177

Coghlan, S., Gyngell, C., & Vears, D. F. (2024). Ethics of artificial intelligence in prenatal and pediatric genomic medicine. *Journal of Community Genetics, 15*(1), 13–24. https://doi.org/10.1007/s12687-023-00678-4

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A. L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, e078378. https://doi.org/10.1136/bmj-2023-078378

Cooperstein, I. B., Marwaha, S., Ward, A., Kobren, S. N., Carter, J. N., Undiagnosed Diseases Network, Quinlan, A., Elkadri, A., Vanderver, A., Rebelo, A., Beggs, A. H., La Spada, A. R., Huang, A., Paul, A., Miller, A., Ward, A., Bale, A., McConkie-Rosell, A., Krokosky, A., … Marth, G. T. (2025). An optimized variant prioritization process for rare disease diagnostics: Recommendations for Exomiser and Genomiser. *Genome Medicine, 17*(1), 127. https://doi.org/10.1186/s13073-025-01546-1

Cornthwaite, M., Turner, K., Armstrong, L., Boerkoel, C. F., Chang, C., Lehman, A., Nikkel, S. M., Patel, M. S., Van Allen, M., & Langlois, S. (2022). Impact of variation in practice in the prenatal reporting of variants of uncertain significance by commercial laboratories: Need for greater adherence to published guidelines. *Prenatal Diagnosis, 42*(12), 1514–1524. https://doi.org/10.1002/pd.6232

De La Vega, F. M., Chowdhury, S., Moore, B., Frise, E., McCarthy, J., Hernandez, E. J., Wong, T., James, K., Guidugli, L., Agrawal, P. B., Genetti, C. A., Brownstein, C. A., Beggs, A. H., Löscher, B. S., Franke, A., Boone, B., Levy, S. E., Õunap, K., Pajusalu, S., … Kingsmore, S. F. (2021). Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Medicine, 13*(1), 153. https://doi.org/10.1186/s13073-021-00965-0

de Koning, M. A., Srebniak, M. I., Oldekamp, E. J., Hahn, D., Diderich, K. E. M., Bruggenwirth, H. T., Santen, G. W. E., Hoffer, M. J. V., & Suerink, M. (2025). Prenatal variants of uncertain significance (VUS): To report or not to report? *European Journal of Human Genetics, 33*(10), 1300–1308. https://doi.org/10.1038/s41431-025-01924-8

Drexler, K. A., Talati, A. N., Gilmore, K. L., Veazey, R. V., Powell, B. C., Weck, K. E., Davis, E. E., & Vora, N. L. (2023). Association of deep phenotyping with diagnostic yield of prenatal exome sequencing for fetal brain abnormalities. *Genetics in Medicine, 25*(10), 100915. https://doi.org/10.1016/j.gim.2023.100915

Dungan, J. S., Klugman, S., Darilek, S., Malinowski, J., Akkari, Y. M. N., Monaghan, K. G., Erwin, A., & Best, R. G. (2023). Noninvasive prenatal screening (NIPS) for fetal chromosome abnormalities in a general-risk population: An evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine, 25*(2), 100336. https://doi.org/10.1016/j.gim.2022.11.004

El-Dessouky, S. H., Sharaf-Eldin, W. E., Aboulghar, M. M., Mousa, H. A., Zaki, M. S., Maroofian, R., Senousy, S. M., Eid, M. M., Gaafar, H. M., Ebrashy, A., Shikhah, A. Z., Abdelfattah, A. N., Ezz-Elarab, A., Ateya, M. I., Hosny, A., Youssef, M. A., Abdella, R., Issa, M. Y., Matsa, L. S., … Abdalla, E. M. (2025). Integrating prenatal exome sequencing and ultrasonographic fetal phenotyping for assessment of congenital malformations: High molecular diagnostic yield and novel phenotypic expansions in a consanguineous cohort. *Clinical Genetics, 108*(1), 33–48. https://doi.org/10.1111/cge.14712

Feuer, O., Holmes, K., Kane, S., Curry, K. M., Ma, D., Chatwin, C. A., Chuldzhyan, M., Quinn, E., & Gorman, N. (2025). AI-scending the scope: Perspectives on the integration and utilization of artificial intelligence and machine learning in genetic counseling graduate programs. *Journal of Genetic Counseling, 34*(4), e70065. https://doi.org/10.1002/jgc4.70065

Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine, 385*(3), 283–286. https://doi.org/10.1056/NEJMc2104626

Gao, Z., Liu, M., Jiang, J., Yang, X., Li, Y., Zhang, Y., Wang, Y., Hua, C., Liu, N., Zhu, X., & Kong, X. (2026). Whole-genome sequencing for the prenatal evaluation of fetal structural anomalies: A prospective multicenter study. *American Journal of Obstetrics and Gynecology, 234*(1), 230–249. https://doi.org/10.1016/j.ajog.2025.08.025

Gargano, M. A., Matentzoglu, N., Coleman, B., Addo-Lartey, E. B., Anagnostopoulos, A. V., Anderton, J., Avillach, P., Bagley, A. M., Bakštein, E., Balhoff, J. P., Baynam, G., Bello, S. M., Berk, M., Bertram, H., Bishop, S., Blau, H., Bodenstein, D. F., Botas, P., Boztug, K., … Robinson, P. N. (2024). The Human Phenotype Ontology in 2024: Phenotypes around the world. *Nucleic Acids Research, 52*(D1), D1333–D1346. https://doi.org/10.1093/nar/gkad1005

Gazdarica, J., Hekel, R., Budis, J., Kucharik, M., Duris, F., Radvanszky, J., Turna, J., & Szemes, T. (2019). Combination of fetal fraction estimators based on fragment lengths and fragment counts in non-invasive prenatal testing. *International Journal of Molecular Sciences, 20*(16), 3959. https://doi.org/10.3390/ijms20163959

Gibbs, A., Braham, R., Ramachandran, V., Roberts, R., Willison, C., Ashraf, T., Cobben, J., Gardham, A., Holder-Espinasse, M., Homfray, T., Mehta, S. G., Tapon, D., Vasudevan, P., & Chandler, N. J. (2026). Is there potential clinical utility in reporting variants of uncertain significance from prenatal sequencing? *Prenatal Diagnosis*. Advance online publication. https://doi.org/10.1002/pd.70192

Gil, M. M., Accurti, V., Santacruz, B., Plana, M. N., & Nicolaides, K. H. (2017). Analysis of cell-free DNA in maternal blood in screening for aneuploidies: Updated meta-analysis. *Ultrasound in Obstetrics & Gynecology, 50*(3), 302–314. https://doi.org/10.1002/uog.17484

Goldlust, I. S., & Bianchi, D. W. (2025). Incidental detection of maternal cancer following cell-free DNA screening for fetal aneuploidies. *Clinical Chemistry, 71*(1), 61–68. https://doi.org/10.1093/clinchem/hvae170

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science, 339*(6117), 321–324. https://doi.org/10.1126/science.1229566

Hanson, B., Paternoster, B., Povarnitsyn, N., Scotchman, E., Chitty, L., & Chandler, N. (2023). Non-invasive prenatal diagnosis (NIPD): Current and emerging technologies. *Extracellular Vesicles and Circulating Nucleic Acids, 4*(1), 3–26. https://doi.org/10.20517/evcna.2022.44

Hanson, B., Scotchman, E., Chitty, L. S., & Chandler, N. J. (2022). Non-invasive prenatal diagnosis (NIPD): How analysis of cell-free DNA in maternal plasma has changed prenatal diagnosis for monogenic disorders. *Clinical Science, 136*(22), 1615–1629. https://doi.org/10.1042/CS20210380

Héron, D., Gerasimenko, A., Frugère, L., Ducourneau, J., Rossi, C., Nava, C., de Sainte-Agathe, J. M., Mignot, C., Lehalle, D., Grotto, S., El-Khattabi, L., Nguyen, T., Garel, C., Blondiaux, E., Milh, M., Desnous, B., Girard, N., des Portes, V., Guibaud, L., … Heide, S. (2025). A large cohort study of prenatal exome sequencing redefines diagnosis in fetal corpus callosum anomalies. *Brain, 148*(12), 4253–4258. https://doi.org/10.1093/brain/awaf311

Horgan, R., Nehme, L., & Abuhamad, A. (2023). Artificial intelligence in obstetric ultrasound: A scoping review. *Prenatal Diagnosis, 43*(9), 1176–1219. https://doi.org/10.1002/pd.6411

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., & Farh, K. K. H. (2019). Predicting splicing from primary sequence with deep learning. *Cell, 176*(3), 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Jeon, S., Lee, S. A., Chung, H. S., Yun, J. Y., Park, E. A., So, M. K., & Huh, J. (2025). Evaluating the use of generative artificial intelligence to support genetic counseling for rare diseases. *Diagnostics, 15*(6), 672. https://doi.org/10.3390/diagnostics15060672

Jin, X., Xu, X., Zhou, A., Zhu, H., Li, Q., Liu, S., Liu, S., Huang, S., Zhang, J., Wang, T., & Liu, H. (2024). Advances in using non-invasive prenatal testing to study genomics related to maternity. *Cell Genomics, 4*(10), 100677. https://doi.org/10.1016/j.xgen.2024.100677

Ren, J., Zhang, Z., Wu, Y., Wang, J., & Liu, Y. (2024). A comprehensive review of deep learning-based variant calling methods. *Briefings in Functional Genomics, 23*(4), 303–313. https://doi.org/10.1093/bfgp/elae003

Kelly, K. E., Galloway, S., Demers, A., Bergner, A. L., & Giordano, J. L. (2026). Genome sequencing for all pregnant persons: Navigating the next frontier in prenatal diagnosis through patient reflections. *Prenatal Diagnosis, 46*(5–6), 719–726. https://doi.org/10.1002/pd.6884

Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., Gammill, H. S., Rubens, C. E., Santillan, D. A., Murray, J. C., Tabor, H. K., Bamshad, M. J., Eichler, E. E., & Shendure, J. (2012). Noninvasive whole-genome sequencing of a human fetus. *Science Translational Medicine, 4*(137), 137ra76. https://doi.org/10.1126/scitranslmed.3004323

Krenn, M., Schmidt, A., Wagner, M., Ernst, M., Graf, E., Zulehner, G., Cetin, H., Zimprich, F., & Rath, J. (2026). AlphaMissense prediction for the evaluation of missense variants in the diagnostic setting of neuromuscular disorders. *Journal of Neuromuscular Diseases, 13*(4), 769–773. https://doi.org/10.1177/22143602251370957

Lee, J., Lee, S. M., Ahn, J. M., Lee, T. R., Kim, W., Cho, E. H., & Ki, C. S. (2022). Development and performance evaluation of an artificial intelligence algorithm using cell-free DNA fragment distance for non-invasive prenatal testing (aiD-NIPT). *Frontiers in Genetics, 13*, 999587. https://doi.org/10.3389/fgene.2022.999587

Lindquist, M., Darrah, S., Stafie, S. T., & Mustafi, D. (2026). Benchmarking AlphaMissense against ClinVar for diagnostic interpretation of missense variants in inherited retinal diseases. *Ophthalmology Science, 6*(2), 100997. https://doi.org/10.1016/j.xops.2025.100997

Liscovitch-Brauer, N., Mesika, R., Rabinowitz, T., Volkov, H., Grad, M., Matar, R. T., Basel-Salmon, L., Tadmor, O., Beker, A., & Shomron, N. (2024). Machine learning-enhanced noninvasive prenatal testing of monogenic disorders. *Prenatal Diagnosis, 44*(9), 1024–1032. https://doi.org/10.1002/pd.6570

Lo, Y. M. D., Chan, K. C. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M. F., Zheng, Y. W., Leung, T. Y., Lau, T. K., Cantor, C. R., & Chiu, R. W. K. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Translational Medicine, 2*(61), 61ra91. https://doi.org/10.1126/scitranslmed.3001720

Lo, Y. M. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet, 350*(9076), 485–487. https://doi.org/10.1016/S0140-6736(97)02174-0

Lord, J., McMullan, D. J., Eberhardt, R. Y., Rinck, G., Hamilton, S. J., Quinlan-Jones, E., Prigmore, E., Keelagher, R., Best, S. K., Carey, G. K., Mellis, R., Robart, S., Berry, I. R., Chandler, K. E., Cilliers, D., Cresswell, L., Edwards, S. L., Gardiner, C., Henderson, A., … Wilson, E. (2019). Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): A cohort study. *The Lancet, 393*(10173), 747–757. https://doi.org/10.1016/S0140-6736(18)31940-8

Mello, M. M., & Cohen, I. G. (2025). Regulation of health and health care artificial intelligence. *JAMA, 333*(20), 1769–1770. https://doi.org/10.1001/jama.2025.3308

Monaghan, K. G., Leach, N. T., Pekarek, D., Prasad, P., & Rose, N. C. (2020). The use of fetal exome sequencing in prenatal diagnosis: A points to consider document of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine, 22*(4), 675–680. https://doi.org/10.1038/s41436-019-0731-7

Mone, F., Abu Subieh, H., Doyle, S., Hamilton, S., McMullan, D. J., Allen, S., Marton, T., Williams, D., & Kilby, M. D. (2022). Evolving fetal phenotypes and clinical impact of progressive prenatal exome sequencing pathways: Cohort study. *Ultrasound in Obstetrics & Gynecology, 59*(6), 723–730. https://doi.org/10.1002/uog.24842

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Petrovski, S., Aggarwal, V., Giordano, J. L., Stosic, M., Wou, K., Bier, L., Spiegel, E., Brennan, K., Stong, N., Jobanputra, V., Ren, Z., Zhu, X., Mebane, C., Nahum, O., Wang, Q., Kamalakaran, S., Malone, C., Anyane-Yeboa, K., Miller, R., … Wapner, R. J. (2019). Whole-exome sequencing in the evaluation of fetal structural anomalies: A prospective cohort study. *The Lancet, 393*(10173), 758–767. https://doi.org/10.1016/S0140-6736(18)32042-7

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology, 36*(10), 983–987. https://doi.org/10.1038/nbt.4235

Qi, Y., Hu, C., Yang, J., Gao, Y., & Yin, A. (2024). Self-learning neural network as a prediction model in non-invasive prenatal testing to detect fetal SNVs. *Journal of Translational Medicine, 22*(1), 707. https://doi.org/10.1186/s12967-024-05433-y

Ramirez Zegarra, R., & Ghi, T. (2023). Use of artificial intelligence and deep learning in fetal ultrasound imaging. *Ultrasound in Obstetrics & Gynecology, 62*(2), 185–194. https://doi.org/10.1002/uog.26130

Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., Ledbetter, D. H., Maglott, D. R., Martin, C. L., Nussbaum, R. L., Plon, S. E., Ramos, E. M., Sherry, S. T., & Watson, M. S. (2015). ClinGen—The Clinical Genome Resource. *New England Journal of Medicine, 372*(23), 2235–2242. https://doi.org/10.1056/NEJMsr1406261

Reilly, K., Rolnik, D. L., Allen, S., Sotiriadis, A., Ong, S., Sonner, S., Blayney, G. V., Fernando, M., Van Mieghem, T., Borrell, A., Langlois, S., & Mone, F. (2025). Performance of international phenotypic criteria for prenatal exome sequencing: Systematic review and comparative diagnostic accuracy study using historical individual participant data. *Ultrasound in Obstetrics & Gynecology, 66*(3), 282–289. https://doi.org/10.1002/uog.29290

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine, 17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Schwammenthal, Y., Rabinowitz, T., Basel-Salmon, L., Tomashov-Matar, R., & Shomron, N. (2025). Noninvasive fetal genotyping using deep neural networks. *Briefings in Bioinformatics, 26*(1), bbaf067. https://doi.org/10.1093/bib/bbaf067

Shear, M. A., Robinson, P. N., & Sparks, T. N. (2025). Fetal imaging, phenotyping, and genomic testing in modern prenatal diagnosis. *Best Practice & Research Clinical Obstetrics & Gynaecology, 98*, 102575. https://doi.org/10.1016/j.bpobgyn.2024.102575

Shen, H., Yang, M., Liu, J., Chen, K., & Li, X. (2024). Development of a deep learning model for cancer diagnosis by inspecting cell-free DNA end-motifs. *npj Precision Oncology, 8*(1), 160. https://doi.org/10.1038/s41698-024-00635-5

Shreeve, N., Sproule, C., Choy, K. W., Dong, Z., Gajewska-Knapik, K., Kilby, M. D., & Mone, F. (2024). Incremental yield of whole-genome sequencing over chromosomal microarray analysis and exome sequencing for congenital anomalies in prenatal period and infancy: Systematic review and meta-analysis. *Ultrasound in Obstetrics & Gynecology, 63*(1), 15–23. https://doi.org/10.1002/uog.27491

Smith, E. J., Hill, M., Peter, M., Wu, W. H., Mallinson, C., Hardy, S., Chitty, L. S., & Morris, S. (2025). Implementation of a national prenatal exome sequencing service in England: Cost-effectiveness analysis. *BJOG: An International Journal of Obstetrics & Gynaecology, 132*(4), 483–491. https://doi.org/10.1111/1471-0528.18020

Sotiriadis, A., Demertzidou, E., Ververi, A., Tsakmaki, E., Chatzakis, C., & Mone, F. (2025). Incremental yield of exome sequencing over standard prenatal testing in structurally normal fetuses: Systematic review and meta-analysis. *Ultrasound in Obstetrics & Gynecology, 65*(5), 552–559. https://doi.org/10.1002/uog.29195

Strauch, Y., Lord, J., Niranjan, M., & Baralle, D. (2022). CI-SpliceAI—Improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. *PLOS ONE, 17*(6), e0269159. https://doi.org/10.1371/journal.pone.0269159

Sun, Y., Wang, X., Jia, Y., Imoto, S., Li, F., Li, C., & Song, J. (2026). Comprehensive review and assessment of multi-species splicing variant prediction: Task-specific deep learning models and genomic foundation models. *Briefings in Bioinformatics, 27*(3), bbag329. https://doi.org/10.1093/bib/bbag329

Swanson, K., Hodoglugil, U., Sparks, T. N., Lianoglou, B. R., Slavotinek, A. M., & Norton, M. E. (2026). Diagnostic yield after postnatal reanalysis of prenatal exome sequencing results. *Prenatal Diagnosis, 46*(5–6), 924–932. https://doi.org/10.1002/pd.6886

Thauvin-Robinet, C., Garde, A., Favier, M., Delanne, J., Racine, C., Rousseau, T., Nambot, S., Bruel, A. L., Moutton, S., Quelin, C., Colson, C., Brehin, A. C., Guerrot, A. M., Rooryck, C., Putoux, A., Blanchet, P., Odent, S., Schaefer, E., Boute, O., … Bourgon, N. (2025). Reanalysis of unsolved prenatal exome sequencing for structural defects: Diagnostic yield and contribution of postnatal/postmortem features. *European Journal of Human Genetics, 33*(5), 675–682. https://doi.org/10.1038/s41431-025-01823-y

Tsui, W. H. A., Ding, S. C., Jiang, P., & Lo, Y. M. D. (2025). Artificial intelligence and machine learning in cell-free-DNA-based diagnostics. *Genome Research, 35*(1), 1–19. https://doi.org/10.1101/gr.278413.123

Turley, P., Meyer, M. N., Wang, N., Cesarini, D., Hammonds, E., Martin, A. R., Neale, B. M., Rehm, H. L., Wilkins-Haug, L., Benjamin, D. J., Hyman, S., Laibson, D., & Visscher, P. M. (2021). Problems with using polygenic scores to select embryos. *New England Journal of Medicine, 385*(1), 78–86. https://doi.org/10.1056/NEJMsr2105065

Van den Veyver, I. B., Chandler, N., Wilkins-Haug, L. E., Wapner, R. J., & Chitty, L. S. (2022). International Society for Prenatal Diagnosis updated position statement on the use of genome-wide sequencing for prenatal diagnosis. *Prenatal Diagnosis, 42*(6), 796–803. https://doi.org/10.1002/pd.6157

Walton, H., Daniel, M., Peter, M., McInnes-Dean, H., Mellis, R., Allen, S., Fulop, N. J., Chitty, L. S., & Hill, M. (2025). Evaluating the implementation of the rapid prenatal exome sequencing service in England. *Public Health Genomics, 28*(1), 34–52. https://doi.org/10.1159/000543104

Xue, Z., Zhou, A., Zhu, X., Li, L., Zhu, H., Jin, X., & Wang, J. (2024). NIPT-PG: Empowering non-invasive prenatal testing to learn from population genomics through an incremental pan-genomic approach. *Briefings in Bioinformatics, 25*(4), bbae266. https://doi.org/10.1093/bib/bbae266

Zemet, R., Parobek, C. M., Adams, A. D., Maktabi, M. A., Shay, L., Meng, L., Liu, P., Dai, H., Xia, F., Eng, C., Van den Veyver, I. B., & Vossaert, L. (2025). Diagnostic yield of exome sequencing for pregnancies with and without fetal anomalies and for stillbirth. *Prenatal Diagnosis, 45*(10), 1313–1324. https://doi.org/10.1002/pd.6817

Zhang, L., Hua, R., Wu, Y., Han, X., Wu, Y., Fei, H., Zhao, X., Chang, C., Gao, L., Chen, Y., Xu, H., Li, N., Yang, J., Wang, Y., Wang, J., & Li, S. (2026). Universal noninvasive prenatal diagnosis for monogenic disorders using cell-free plasma DNA. *Genome Medicine, 18*(1), 4. https://doi.org/10.1186/s13073-025-01588-5

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/164501