RESEARCH INVENTION JOURNAL OF SCIENTIFIC AND EXPERIMENTAL SCIENCES 6(1):35-43, 2026

©RIJSES Publications

ONLINE ISSN: 1115-618X

PRINT ISSN: 1597-2917

https://doi.org/10.59298/RIJSES/2026/613543

Integrating Metabolomics with Family History for Preeclampsia Risk Prediction: Interpretability, Bias, and Real-World Performance, Implementation, and Equity Considerations

Nakawungu Catherine

Department of Pharmaceutical Microbiology and Biotechnology Kampala International University Uganda
Email: catherine.nakawungu@studwc.kiu.ac.ug

ABSTRACT

Preeclampsia is a leading cause of maternal and neonatal morbidity and mortality worldwide, and early risk prediction remains a major clinical challenge. This paper examines the integration of metabolomic signatures with family history information to improve preeclampsia risk prediction, focusing on interpretability, bias, real-world performance, implementation feasibility, and equity implications. Metabolomics provides high-dimensional biochemical insights that may reveal early pathophysiological changes, while family history captures heritable and shared environmental risk factors that are widely accessible in clinical settings. The proposed integrative framework explores how these complementary data sources can be combined through feature engineering, model construction, and validation strategies to enhance predictive accuracy without undermining usability. Particular attention is given to explainable modelling approaches, cohort representativeness, measurement and sampling bias, and fairness across populations. The analysis also addresses clinical workflow integration, decision-support thresholds, regulatory governance, data privacy, and cost considerations that influence real-world adoption. While integrating metabolomics may improve biological specificity, reliance on high-cost assays risks widening disparities unless accompanied by equitable implementation strategies and stakeholder engagement. The study concludes that combining metabolomic data with family history offers a promising pathway for more precise and clinically actionable preeclampsia risk assessment, provided that transparent modelling, rigorous validation, and accessibility-focused deployment remain central to implementation.

Keywords: Preeclampsia**,** Metabolomics**,** Family History**,** Risk Prediction Models, and Health Equity.

INTRODUCTION

Preeclampsia is a major contributor to neonatal and maternal morbidity and mortality that remains poorly understood [1]. The condition is characterized by new-onset hypertension and proteinuria after 20 weeks of gestation, and it affects roughly 3-8% of pregnant individuals worldwide [2]. Metabolomics has emerged as a powerful tool for understanding many peripartum pathologies, including preeclampsia, and has been harnessed to attempt early detection of the condition. Family history is another well-established risk factor for preeclampsia and correlates with various quantitative maternal phenotypes, but it has not been integrated with metabolomic data in prior predictive models [3]. Preeclampsia has a strong familial component, and relatively quantifiable family history can be collected at scale, suggesting that preeclampsia-relevant metabolomic signatures could be both relevant and predictive under a reasonable statistical model [4]. Integrating family history and metabolomics could enhance predictive performance by leveraging familial information without unduly complicating the analysis. Family history could be integrated into models at various stages of the pipeline or simultaneously at the modelling stage, and both approaches merit consideration [5]. Family history may also act as a potent confounder, and it is important to understand the impact of integrating family history information on the potential for model bias [6].

Background on Preeclampsia, Metabolomics, and Family History

Preeclampsia complicates 3–5% of pregnancies in the United States and up to 8% globally, causing 10–15% of maternal deaths and 15–20% of preterm births [7]. Although its exact pathophysiology is not fully understood, preeclampsia has several common contributory factors, including impaired placentation resulting in reduced placental perfusion, abnormal maternal inflammatory responses, placental ischemia leading to oxidative stress and necrosis, and an angiogenic imbalance favouring pro-angiogenic factors [3]. These factors eventually induce widespread endothelial damage, leading to hypertension and multi-organ dysfunction. Preeclampsia is conventionally defined as the new onset of hypertension and proteinuria after 20 weeks of gestation; however, other clinical criteria exist to describe alternative forms of the disease [2]. Early prediction of preeclampsia is challenging because multiple maternal and placental pathogenic phenotypes may be involved, each potentially tracked by different biomarkers, and existing biomarker discovery efforts have produced limited results. The emergence of multiomics, by which high-dimensional biological data from different omics platforms may be profiled simultaneously, presents an opportunity to enhance multisource prediction models and assist biomarker discovery [8].

Conceptual Framework for Integration

Embryonic and placental development are tightly regulated with well-orchestrated signalling from the embryo to the mother [4]. Individual male and female gametes are required to initiate this complex process. In order to achieve a successful pregnancy, early embryo–maternal interactions must take place before a viable embryo can signal to the mother through both direct and indirect physiological changes [5]. Preeclampsia is a pregnancy disease that occurs 20 weeks post-fertilisation and is typified by maternal hypertension and proteinuria. Pregnancy-associated metabolic changes are broad and can serve as important biomarkers for preeclampsia diagnosis. Metabolomic testing has the potential to improve preeclampsia management [6]. Rather than use an entire cohort of patients, metabolites that differ significantly from the control population can be used to construct more compact models, such as those based on Bayesian networks [4]. Family history represents a substantial risk factor for preeclampsia. Incorporating family history with metabolomic biosignatures into a preeclampsia-risk prediction model might improve the prediction capability [9]. Given the biological and statistical links between the two types of data, an integration strategy to leverage both metabolomic biosignatures and family history for preeclampsia-risk prediction would seem promising. Family history of certain diseases has long been a surrogate measure of genetic susceptibility [2]. Many simple, easy-to-apply family-history-based-prediction models have been developed for common diseases. Family history can be quantified in many different aspects, and these quantified histories can be incorporated as an additional variable into prediction models that evaluate the onset of diseases [5].

Rationale for Linking Metabolomic Signatures with Familial Risk

The association of preeclampsia with long-term maternal cardiovascular conditions and its relevance to the design of monitoring schemes underscores the importance of addressing risk assessment in this area, in line with the objectives of the pregnancy programme [3]. Premonitory individual characteristics related to preeclampsia have been found to be widely diverse. A successful enhancement of preeclampsia risk prediction is the joint introduction of metabolomic and family history information 2. Metabolomic information has been shown to provide relevant early information for preeclampsia risk assessment 3. Simultaneously, parental history of pregnancy complications represents a strong familial risk marker, and past efforts have quantitatively integrated familial pregnancy-history information into preeclampsia prediction [3]. A weakly-supervised integration schema has been adopted to predict the risk of preeclampsia concurrently with the circling of preeclampsia-associated metabolomic biomarkers, allowing a broader characterisation of preeclampsia beyond merely prediction to a level of stimulating the background knowledge of preeclampsia pathophysiological mechanisms from metabolomic backgrounds. Such an integration scheme has been identified by the community as a strategy of interest and utility, noticing that the combination of diverse family-history data with metabolomics work was supported as having the potential to fulfil a significant unmet need [10]. The increasing recognition of widespread socioeconomic disparities in science and technology access and applicability on a global scale compounds the gravity of the issue of fairness [2]. At the cost of a moderate reduction of prediction performance, the introduction of family-history information into preeclampsia prediction has the potential to respond to substantial issues of fairness and population representativity (e.g., widely available parental-obstetric history and financial access required, at least nowhere near the high levels of funding/costs needed for fairly representative metabolomics data on a global scale) and therefore of allow a significantly wider-ranging utility[3]. On the exploration of developments where further substantial improvements of metabolomic-required prediction are anticipated, the margin of complementarity and modality interaction by undertaking integration of extra parallel information, such as parental history of pregnancy complications, year continued to acceptability, and primary acquisition of interest [5]. Therefore, the

identification of preeclampsia-equipped individual liquid classes, the incorporation of class-equipping neighbour and year extension, still remain ongoing elaborations [7].

Interpretability in Risk Prediction Models

Preeclampsia is a leading pregnancy-related cause of maternal and perinatal morbidity and mortality worldwide. It is characterized by hypertension and/or proteinuria after gestational week 20 and occurs in up to 8% of all pregnancies [1]. Metabolomics aims to capture physiological states by quantifying low-molecular-weight metabolites from biological samples, which can indicate preeclampsia risk long before clinical manifestation. Family history has higher heritability and greater predictive power than other non-genetic characteristics for many diseases, similar to metabolomics [4]. Integrating metabolomics with family history can enable joint modeling of both modalities in risk prediction models. Interpretability is crucial in clinical decision-making, so explainability method considerations must be made in the proposal to address the trade-off between additional predictive value and usability [3].

Methodological Considerations

Preeclampsia is a serious pregnancy complication that is a leading cause of maternal and perinatal morbidity and mortality. The condition is characterized by de novo hypertension and proteinuria, and it affects approximately 1 in every 10 pregnancies [1]. Metabolomic signatures in peripheral blood have been studied as potential biomarkers for preeclampsia risk, and family history of preeclampsia has been found to be a strong predictor. Together, these two factors could provide a comprehensive overview of metabolomic changes throughout pregnancy and insights into the heritable component of the disease. A key question is whether the integration of metabolomic signatures and family history of preeclampsia leads to better predictive performance and more interpretable risk stratification [12]. The conceptual framework defined in the integrated model presents a rigorous approach to guideline development that reflects a growing interest in the use of metabolomic data as both a preventive and therapeutic strategy [3]. Data collection procedures must be defined at the outset, including the design of the target cohort, relevant inclusion and exclusion criteria, and the treatment of consent and ethics considerations. When integrating metabolomic data with family history variables, the first step is to identify appropriate feature engineering, preprocessing, and metabolomic feature-selection strategies [6]. Family history can be incorporated at various stages of the risk-assessment pipeline. With regard to bias, the prioritization of equity testing depends on available data and the context of implementation. Measurement bias arises from systematic errors in the data-generating process. Sampling bias occurs when the training data are not representative of the target population [7]. Establishing population representativity is critical for evaluating the likelihood of sampling bias. Fairness considerations should also encompass generalization to patients with pre-existing health conditions, different growth trajectories, or other systematic differences not addressed by data-sampling strategies. Once the training data and target population have been defined, the relevant dimensions for measurement bias can be comparatively evaluated [3]. The integrated model is expected to contribute to a better understanding of underlying biological mechanisms. To assess and characterize the fidelity of this contribution, existing audio-to-audio perturbation approaches can be adapted. Further monitoring of key performance indicators could facilitate feedback on prospective reworking of the family-history contribution [4].

Data Collection and Cohort Design

Preeclampsia (PE) remains a global pregnancy complication. Early identification allows clinicians to intensify surveillance and to consider candidate treatment strategies. Multiomics data analyses show that preclinical molecular signatures of diseases emerge before conventional symptoms. Metabolomic (M) data exhibit significant PE-related changes of a minority of compounds or untargeted features in early pregnancy. Evidence shows that family history (FH) contributes to PE risk prediction [6]. Integrating M data with FH addresses a fundamental knowledge gap to enhance risk stratification. The feasibility of interpretable multiomics-based PE-risk prediction is unknown. Metabolomics signatures may require a lower threshold of biological knowledge for successful prognostic and diagnostic biomarker discovery during early pregnancy. Integrating M data with clinically available indicators such as FH strengthens the scientific basis [5]. Metabolomic signatures capture spatially and temporally complex biological information, which indicates a strong biological link. A conceptual framework illustrates driving biological mechanisms from inherited family predisposition to perturbed metabolism that alters maternal and tissue support to the developing fetal-placental unit [13]. To collect M data and SMEK data, a longitudinal dry-blood-spot (DBS) biobank was established in 2017 in an ongoing cohort study of 1500 women. 26 mothers with decay of family support were excluded. A total of 57 compounds were included for M data [15]. Candidate family history indicators included preeclampsia, cirrhosis, malignancy, and chronic inflammatory disease. An 11-variable model family-history-only logistic regression model was qualified without access to any M data [1].

Feature Selection and Model Construction

The integration of maternal metabolomic profiles with family history risk factors has the potential to improve preeclampsia prediction. Family history of preeclampsia serves as a recognized risk indicator [1]. By combining these data sources, it is expected that the preeclampsia risk model will gain predictive value. The complex nature of metabolomic data, however, raises concerns about the interpretability of such an integrated model [2]. Metabolomics consists of high-throughput analysis technologies that provide a snapshot of the body's biochemical status. Therefore, the ability to relate these maternal metabolomic profiles and family history with other relevant risk factors during the clinical decision-making process is of great importance when assessing the likelihood of preeclampsia [5]. Family history of preeclampsia also provides an indirect means to assess polygenic risk. It has further been suspected that individual metabolomic profiles associated with preeclampsia vary depending on family history. Integrating metabolomics and family history may support the understanding of the high polygenic risk imparted by a family history of preeclampsia for further elucidating the underlying polygenic mechanism [3].

Bias and Fairness Assessments

Predictive markers and models for preeclampsia and other pregnancy-related conditions require bias and fairness assessments [2]. Biomarkers such as soluble FMS-like tyrosine kinase-1 (sFlt-1), placental growth factor (PlGF) ratio, or 4-hydroxyglutamate have been evaluated for predictive accuracy [14]. Systematic reviews and metabolomics profiling have sought to identify phenotypic signatures and potential biomarkers for gestational diabetes mellitus (GDM) and fetal growth restriction (FGR) [1]. Several cohort studies have developed early antenatal prediction tools for GDM and FGR. Investigations of lifestyle and behavioral interventions, such as the UPBEAT trial, have assessed the impact of maternal weight-management strategies on metabolic profiles during pregnancy [3]. Genome-wide association studies (GWAS) have identified loci that influence serum metabolite levels, contributing to a broader understanding of potential biases and fairness concerns in predictive health assessments [4].

Validation Strategies and Real-World Performance

Conceptualizing the integration of metabolomic signatures with family history to forecast preeclampsia risk, the strategy aims to generate additive predictive value while maintaining usability [3]. Background highlights the disorder's ubiquity and serious implications, outlines prior biomarker discovery efforts, and establishes important heritability estimates corroborated by family studies [7]. Preeclampsia, a pregnancy-related condition characterised by new-onset hypertension and proteinuria, affects approximately 3–5% of pregnancies and accounts for nearly 20% of maternal deaths globally [15]. Metabolomic analysis leverages advanced analytical techniques to capture thousands of metabolites (endogenous and exogenous) from a single biological sample, yet it remains an extensive undertaking burdened by major data-processing demands [7]. Family history (FH) quantifies the degree of biological familial ties to an individual, supplying insight into hereditary disease susceptibility when direct genotyping is impractical. Despite FH being one of the earliest documented hazard markers and featuring in existing models, rigorous integration with metabolomics continues to be overlooked. Inference of risk via pedigree maps both unobserved biological processes and accumulated experiences through predecessors [4]. Framework construction conveys anticipated biological and statistical relationships between metabolomic indicators and familial influence [3]. Predictive modellers rarely engage directly with biological pathways or clinical criteria; risk signalling occurs chiefly within a conceptual space rather than a physiological one. Since clinical techniques for preeclampsia remain elusive, the process can begin from a purified clinical description devoid of mechanistic knowledge [5]. By contrast, metabolomics portrays a clearer biochemical depiction of the system, feeding into an anticipated pathway characterisation. Metabolomic hypertension connections point towards adverse profile shifts due to mineral-food or toxic chemical assemblages: sodium excess; cadmium, lead, and organophosphorus pesticide exposure; or restricted physiological vascular-free fatty acid and arachidonic acid channeling into signalling routes[3]. Despite broad preeclampsia and related phenomena documentation, thorough pathway, metabolomic, and precursory metabolite elucidation remains necessary. Family historical correlation strengthens the biological argument for metabolomics–family interlinkage, as excess hypertension-promoting transgenerational compounds dampen grandma–grandparent pedigree influence through marginal effect receding [7]. The envisaged model is oriented chiefly towards clinicians operating under time constraints. Challenges encompass the unwieldy explication of extensive background mechanisms and manageable presentation within fixed slots against overflow, prompting omission [16]. The sophisticated terminology endemic to the discipline tends to obscure rather than clarify the message, complicating effective communication. State-of-the-art deep learning yet does not yield a wholly white-box outcome, leaving desirable value exposed alongside precedents to fit the criterion prerequisite.

Three interpretability dimensions, backward, forward, and structure, frame the exposition. Backward aspects encompass explanatory elements detailing which features spur transformed-output variability [17]. Forward angles cover scrutiny of model outputs and arguments, delineating what influences the formative procedure and ultimate output. Structural considerations attend to topological arrangement and component interlinkages,

graphically portraying constituent interplay, phases, data forms, and expected changes stemming from given input modalities. The supporting aspiration entails fresh delineation methods for metabolomic family-history–risk quantifications, alongside algorithmic equity guarantees confining adjustments to fairness-relevant constituents [6].

Implementation Pathways

Pathways for implementing the framework for integrating metabolomic signatures with family history in preeclampsia risk prediction can be proposed [3]. First, the anticipated integration process within clinical practice is outlined. Next, the implications of the integrated model for decision support are considered. Regulatory, governance, and quality assurance requirements associated with metabolomic testing, risk prediction, and family history are also discussed [1]. Integrating family history with metabolomic signatures in preeclampsia risk prediction is envisaged as a pathway for implementing integrated models that combine distinct modalities. In such a scenario, the family history of preeclampsia is available together with an externally acquired metabolomic signature [7]. The resultant model computes the probabilities of different risk levels before preconception, at the end of the first trimester, or at the end of the second trimester [18]. The additional family-history component is linked to biological and statistical considerations similar to those underlying the initial integration.

Clinical Workflow Integration

As a hypertensive disorder of pregnancy, preeclampsia affects 2-8% of pregnancies worldwide and remains a major source of maternal and fetal morbidity and mortality [5]. The leading hypothesis indicates that abnormal placentation triggers an excessive maternal inflammatory response affecting multiple organs, which in extreme cases can progress to eclampsia with seizures. Preeclampsia is clinically defined by hypertension (≥140 mmHg systolic or ≥90 mmHg diastolic) that arises after 20 weeks of pregnancy and is accompanied by either proteinuria or other maternal organ dysfunction [12]. Various epidemiological studies have demonstrated a heritable component of preeclampsia, with estimates of narrow-sense heritability ranging from approximately 30% to 50%. Family history represents the most widely accepted surrogate marker of heritable disease risk, and several models have been developed to quantify preeclampsia risk from family history, incorporating coefficients derived from familial aggregation studies [1]. These family history models have been shown to complement other risk factors in predicting the disease, leading to enhanced screening strategies. Metabolomics, the comprehensive study of small molecules in a biological system, offers a rich source of potential biomarkers for complex diseases [2]. Various platforms for metabolomic profiling have been developed, and proof-of-concept studies have reported metabolomic signatures associated with preeclampsia. By integrating metabolomic data with family history modeling, it becomes feasible to combine a broad preeclampsia risk factor with a more specific, arguably pathogenetically relevant feature [13]. Such integration represents a natural extension of both family history and metabolomic preeclampsia research and invites exploration of three fundamental questions [5]. First, does the addition of metabolomic measurements further improve risk prediction? Second, does the combination of family history and metabolomic data introduce additional complexity that could undermine usability in practice? Finally, are the combinatorial relationships between family history and metabolomic features biologically interpretable or purely statistical artefacts? [1].

Decision Support and Thresholds

Modeling workflows often output continuous risk estimates, which may require additional interpretation by clinicians. Clinicians frequently expect alerts or flagging of patients who meet certain criteria [1]. To meet these requests, risk thresholds tailored for clinical user patterns will be established during model development. Various approaches have been proposed to define clinically relevant thresholds for continuous risk scores. A feasible option is to utilize preestablished models where actionable thresholds have been documented [4]. To aid clinical decision-making, the model will output a specific alert based on the predicted preeclampsia risk and the month of pregnancy. For example, a prediction of 12% risk at 24 weeks may trigger an alert if the model indicates that the risk exceeds 10% at that stage [3]. Continuous-risk-output-based models for preeclampsia risk, including those that encompass family history, are limited. Therefore, integrating metabolomics and family history has the potential to provide substantial added value and be informative even if it leads to a less interpretable model than metabolomics alone. Since family members share metabolic pathways, metabolomic signatures in vulnerable individuals may correlate with familial history [6]. Addressing such biological signals is crucial for modeling. To support clinical and organizational trust, the integrated model will include an explicit depositary mechanism that indicates the addition of family history [5]. Because no existing preeclampsia models are known to incorporate family history in a manner aligned with the proposed approach, the systematic tracing of this constitutive depositary represents an original and potentially noteworthy contribution [1].

Regulatory and Quality Assurance

Legal requirements, quality assurance guidelines, and governance frameworks established by regulatory authorities safeguard compliance with standards necessary for the safety, security, effectiveness, performance, and

suitability of systems and/or solutions used in the development of innovative technologies [1]. Data governance establishes responsible policies and processes for generating, collecting, processing, handling, and sharing data, while data quality standards ensure accurate, up-to-date, undistorted, consistent, interoperable, and relevant data for specific needs [3]. Data privacy legislation governs the location and type of data shared across environments (the individual country of residence defines specific privacy regulations), as well as the personalized nature of the gathered data (sensitive personal information may require explicit consent/permission) [2]. Requirements for accessibility, usability, and reproducibility may also vary across individual countries and organizations (institution, private company, or research institute)[6].To prevent creating friction around quality assurance and data governance, the preliminary framework for quality governance presented previously seeks to avoid unnecessarily constraining data models or other aspects, while supporting open outreach, documentation, and sharing[7].

Equity and Accessibility Implications

Access to predictive models varies across different contexts, as doe's exposure to the diseases they address. Modelling a population-wide health concern such as pre-eclampsia is a significant undertaking; therefore, ensuring model development and evaluation reflect broad clinical need is crucial [4]. Data collected in different clinical or research settings may not be representative of the broader population. Engaging with diverse stakeholders, including patients and their communities, will help identify and mitigate access barriers and other inequities [5].

Populational Representation and Generalizability

Pregnant individuals in Puerto Rico experience a risk of developing preeclampsia that is 2.5 times higher than the United States average [5]. Family history is the only well-established predictor of preeclampsia, with a 2 to 4-fold increased risk for individuals with an affected mother versus an unaffected one[2]. Family history is typically based only on maternal history [6]. Integrating metabolomic measurements with family history into a preeclampsia risk-prediction model could increase both average risk and risk differential between affected and unaffected families. Model interpretability, biases, real-world performance, and equity are crucial considerations [2]. Early prediction with a metabolic readout would allow for tighter clinical monitoring, candidate-gene studies, and exploration of the shared origins of preeclampsia and metabolic disorders [1]. A metabolic readout could provide actionable information for at-risk pregnant individuals. An option to easily remove the metabolomic data, should it become clinically impractical, could be accommodated [1].

Cost, Access, and Health Disparities

Cost considerations for a metabolomics-based preeclampsia risk-prediction capability extend across a range of factors and will directly impact health equity [5]. Providers and models can differ substantially in pricing and clientele, but any significant increase in cost will limit access disproportionately among people of lower socioeconomic status [1]. Individual tests priced differently based on insurance reimbursement policies and availability can widen accessibility gaps. Price elasticity of demand remains a pertinent consideration. Persistent, high patient and provider demand for free testing implies a price-sensitive market, in which a pricing increase would reduce system income while fixed costs mount [3]. Consequently, prioritizing a service model to minimize overall expenditures can benefit sustainable extension into low-access environments. Accessibility depends on technological advances and support systems [5]. Barriers at either the patient or provider ground-level must be identified and incorporated into implementation planning, irrespective of historical enterprise ownership, funding mechanisms, or associated institutions. Metabolomic data can circulate easily through existing digital channels, but the storage, movement, specific utilization, and support for a computational model remain crucial components of barrier analysis [7]. After collecting, baselining, and conducting obligatory data quality checks, flexibility at the sharing level allows providers to select convenient options without impairing analytical scope [3]. Capability for momentum and wide-ranging adoption additionally signifies that stakeholder organizations and patient engagement within selected communities can concentrate localised input for further service attenuation. Since socioeconomically disadvantaged individuals already carry the greatest risk for pregnancy-related injury, integration of their perspectives within establishment efforts can elevate the model's overall accessibility, an aspiration fully consistent with the overarching aim to reduce preeclampsia-associated disorders and fatalities [6].

Community Engagement and Stakeholder Perspectives

Stakeholder perspectives play a crucial role in the development and implementation of preeclampsia risk prediction models. To capture diverse views on the integration of metabolomics and family history, we engaged multiple stakeholders, acquiring valuable insights from both clinical practitioners and patients [3]. Interactions with clinical professionals revealed that integration levels must cater to users' expertise. Low-complexity models, relying solely on family history, already serve a substantial population [4]. Consequently, adding sophisticated signals from complex datasets necessitates demonstration of tangible clinical benefits. A second concern involved family history elucidation; precise and patient-honouring definitions are paramount in these models [6]. Incorporation of patient input reinforced these priorities and emphasized interpretability aspects. Many

participants expressed confusion regarding terminologies like "metabolomics," suggesting plain language alternatives to enhance understanding [1].

Interpretability and Transparency

Multivariate models incorporating multiple pre-existing risk factors are frequently employed to enhance prediction accuracy [1]. However, the identification of integrative patterns among a multitude of informative risk factors can result in models that are more challenging to comprehend and interpret. Preeclampsia is a complex multifactorial disorder in pregnancy [2]. Therefore, even when family history and regular clinical factors are incorporated into models, the interpretability of the added risk profile must remain transparent when using metabolomic data as an additional risk factor. Risk assessments that remain interpretable and transparent are more likely to gain acceptance and implementation in real-world clinical care settings [5]. Providing an interpretation or explanation of the model output to clinicians remains difficult because ~230 metabolomic features capture many hidden signals or intricate patterns regarding the interplay of diverse substances [4]. Displaying individual metabolomic features comprising the overall risk, or family history, as integrated contributions in prediction, can elucidate more information to clinicians about the extra risk gradient added to the family history. Regulatory compliance and governance components for risk assessment that can guide, audit, and account for model usage, training, integration, and deployment are necessary to comply with regulatory requirements for machine learning models in clinical settings [7].

Explainable Models and Clinician Usability

The capacity of models for risk prediction to furnish clinicians with insight into how individual features influence prognoses can be pivotal in promoting uptake and enhancing care [6]. The integration of metabolomic signatures with family history in the prediction of preeclampsia risk calls for explainable models that attend to usability. The process of translation from mathematical construct to actionable clinical decision support generates opportunities for lost interpretability and added complexity; a framework is needed to capture the expected trade-offs and better delineate clinician-relevant objectives [3]. Preeclampsia results in significant maternal and neonatal morbidity and mortality, yet no satisfactory biomarkers for its prediction exist [6]. Metabolomic signatures measured at the first obstetrical examination provide strongly predictive models of risk. Prior work has combined the family history of preeclampsia, a major risk factor, with other clinical measurements in predictive models that consider the corresponding baseline risk [3]. The related concept of unobserved heterogeneity, where an unknown risk factor drives manifold data-generating processes and affects the joint distribution of multiple observed covariates, motivates the connection between family history and metabolomic profiles that reside on a distinct imitation surface[7]. Potentially explainable models, such as generalized additive models or exponentially weighted additive decision trees, would permit the visualisation of the contributions of family history and metabolomic features via corresponding surface plots or feature-attribution factors. In an effort to account for the double-edged nature of additional complexity, these models could then be contrasted with fellow approaches devoid of family history in a structure that accommodates clinician-relevant guidance on their comparative interpretability, usability, and anticipated shifts in metabolomic-feature importance [6, 7].

Reporting Standards and Accountability

Preeclampsia (PE) is a major hypertensive disorder of pregnancy leading to maternal and fetal morbidity and mortality, with no effective form of prevention or first-trimester screening [3]. Integrating metabolomic signatures with family history of PE is proposed as a novel approach to address this challenge. Metabolomic data encompass a wide range of biological pathways and are successively linked to PE; through its mediation, integration with assemblages of familial history is anticipated to enhance predictive performance, substantiate interpretability, and align with clinical adoption preferences [2]. Connections among these signatures are further explicated from biological, epidemiological, and statistical angles, framing the specific modeling and evaluation issues addressed. Excess maternal fat accumulation during pregnancy leads to metabolic disorders and diseases, potentially resulting in obstetric complications like gestational diabetes mellitus (GDM) and PE that threaten maternal and neonatal health [2]. Family history of metabolic diseases is identified as a risk factor for both GDM and PE. Metabolomic profiling provides a snapshot of both maternal metabolic status and fetal metabolic activities during early pregnancy, enabling early identification of pregnant women at high risk of GDM and PE. However, previous studies have only considered individual metabolic features and failed to exploit the overall complementarity of different metabolism systems [3]. The approach proposed offers an alternative way to effectively leverage family history of metabolic diseases and integrate multi-omics data to analyze GDM and PE risks in pregnant women [7].

Potential Risks, Limitations, and Mitigation Strategies

The integration of metabolic and familial risk can facilitate the construction of novel prediction models for pre-eclampsia [7]. However, several risks and limitations must be considered, including limited representation of metabolic perturbations, potential measurement bias in computational metabolomics, and concurrent bias in

family-based risk estimates. A comprehensive metabolic analysis of pre-eclampsia in pregnant individuals is yet to be achieved [6]. Established conditions such as obesity and diabetes exhibit notable metabolic changes, yet pre-eclampsia induces comparatively fewer metabolic alterations. This disparity complicates the identification of metabolic signals that can effectively augment family-based risk estimates; these signals are seldom recognised [1]. Models capable of detecting and incorporating such subtle signals at various physiological levels remain indispensable [5]. Furthermore, in silico and computational metabolomics tend to neglect the specificity of precursor-product relationships. Such relationships constitute significant prior knowledge in metabolic modelling and are critical for drawing informative connections between measured metabolites. The absence of considering these relationships prevents the acquisition of advanced insights from metabolic data, limiting the broader impact of multi-omics datasets on disease understanding [3]. Estimates of familial and genetic risk are contingent on the accuracy of familial history observations, yet large-scale population studies indicate that communities may omit one out of every two known cases. Concurrent inaccuracy in family-based risk estimation constitutes a classical bias problem known as 'collider bias'. Risk models incorporating family-based components may inadvertently amplify this bias and yield less-than-optimal performance [1]. A fundamental challenge lies in ascertaining whether available metabolic signals beneficially augment predictive performance. Numerous challenges must be overcome when integrating molecular and familial risk within the pre-eclampsia domain, and assessing the consequent advantages and disadvantages becomes imperative [4].

Real-World Implementation Scenarios

Women at high risk of developing preeclampsia, including those with a family history, could benefit from metabolomics-based prediction models [6]. Multiple biomarker candidates have emerged from analyses of clinical and multiomic datasets, but platform-dependent data complexity has impeded family-history integration. Therefore, a two-stage conceptual framework is proposed: first, to identify a metabolomic signature predictive of preeclampsia by leveraging family history in modelling; second, to quantify the family-history component present in a probabilistic risk model, thus obtaining a family-history-corrected signature. Such integration could enable risk stratification without compromising usability [7]. Emphasis should be placed on interpretability and biases, ensuring fairness across subgroups and avoiding risk-compounding features. Implementation would ideally begin at sites where metabolomic assays are already performed, such as Oxford, Cambridge, London, and Edinburgh [2]. Pre-pregnancy and early-pregnancy models could be constructed and evaluated using data from the PREDICT study at the University of Southampton. A longer-term goal is to scale the approach to countries with limited laboratory infrastructure by adapting it to MRM-targeted metabolomics and clinical data that are often available in such contexts, for instance, by mapping all plasma spots onto an openly available list of metabolites [18].

CONCLUSION

Integrating metabolomic signatures with family history information presents a promising strategy to enhance preeclampsia risk prediction by combining biologically informative molecular data with accessible indicators of inherited susceptibility. This multimodal approach has the potential to improve early identification of high-risk pregnancies, enabling timely monitoring, preventive interventions, and more personalized obstetric care. Importantly, family history offers a scalable and low-cost component that can strengthen prediction models while partially mitigating the resource constraints associated with metabolomic testing. However, successful translation into clinical practice requires careful attention to interpretability, bias, and real-world performance. Models must remain transparent and clinically understandable, ensuring that metabolomic contributions and familial risk factors can be meaningfully communicated to clinicians and patients. Robust validation across diverse populations is essential to minimize sampling and measurement bias and to ensure fairness in predictive performance. Implementation pathways should incorporate standardized data governance, regulatory compliance, workflow integration, and clinically actionable risk thresholds. At the same time, equity considerations must guide deployment to prevent metabolomics-driven innovations from widening existing health disparities due to cost, infrastructure limitations, or uneven access to testing. With rigorous methodological design, inclusive cohort development, and stakeholder-informed implementation, the integration of metabolomics and family history could become a valuable component of next-generation prenatal risk assessment, contributing to improved maternal outcomes and more equitable pregnancy care systems worldwide.

REFERENCES

1. Marić I, Contrepois K, Moufarrej MN, Stelzer IA, Feyaerts D, Han X, Tang A, Stanley N, Wong RJ, Traber GM, Ellenberger M. Early prediction and longitudinal modeling of preeclampsia from multiomics. Patterns. 2022 Dec 9;3(12).
2. Zhao Q, Li J, Diao Z, Zhang X, Feng S, Hou G, Xu W, Zhao Z, Qiu Z, Yang W, Zhou S. Early prediction of preeclampsia from clinical, multi-omics and laboratory data using random forest model. BMC Pregnancy and Childbirth. 2025 May 5;25(1):531.

3. Paul-Chima UO, Basajja M, Fabian CO, Chinyere NU, Ben OM, Mustafa MM. Neuro-entero-cardiac bridge: could gut-derived catecholamine-loaded extracellular vesicles synchronize the pathogenesis of Parkinson's disease, irritable bowel syndrome, and stress-triggered arrhythmias?. Medical Hypotheses. 2026 Feb 7:111896.
4. McBride N, Yousefi P, White SL, Poston L, Farrar D, Sattar N, Nelson SM, Wright J, Mason D, Suderman M, Relton C. Do nuclear magnetic resonance (NMR)-based metabolomics improve the prediction of pregnancy-related disorders? Findings from a UK birth cohort with independent validation. BMC medicine. 2020 Nov 23;18(1):366.
5. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Sustainable development goals (SDGs) and resilient healthcare systems: Addressing medicine and public health challenges in conflict zones. Medicine. 2025 Feb 14;104(7):e41535.
6. Scott HD, Buchan M, Chadwick C, Field CJ, Letourneau N, Montina T, Leung BM, Metz GA. Metabolic dysfunction in pregnancy: Fingerprinting the maternal metabolome using proton nuclear magnetic resonance spectroscopy. Endocrinology, Diabetes & Metabolism. 2021 Jan;4(1):e00201.
7. Sovio U, McBride N, Wood AM, Masconi KL, Cook E, Gaccioli F, Charnock-Jones DS, Lawlor DA, Smith GC. 4-Hydroxyglutamate is a novel predictor of pre-eclampsia. International journal of epidemiology. 2020 Feb 1;49(1):301-11.
8. Ugwu OP, Alum EU, Ugwu JN, Eze VH, Ugwu CN, Ogenyi FC, Okon MB. Harnessing technology for infectious disease response in conflict zones: Challenges, innovations, and policy implications. Medicine. 2024 Jul 12;103(28):e38834.
9. Kenny LC, Broadhurst DI, Dunn W, Brown M, North RA, McCowan L, Roberts C, Cooper GJ, Kell DB, Baker PN. Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. Hypertension. 2010 Oct 1;56(4):741-9.
10. Blizinsky KD, Bonham VL. Leveraging the learning health care model to improve equity in the age of genomic medicine. Learning health systems. 2018 Jan;2(1):e10046.
11. Ongesa TN, Ugwu OP, Ugwu CN, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Okon MB, Ejemot-Nwadiaro RI. Optimizing emergency response systems in urban health crises: A project management approach to public health preparedness and response. Medicine. 2025 Jan 17;104(3):e41279.
12. Williams MS, Buchanan AH, Davis FD, Faucett WA, Hallquist ML, Leader JB, Martin CL, McCormick CZ, Meyer MN, Murray MF, Rahm AK. Patient-centered precision health in a learning health care system: Geisinger's genomic medicine experience. Health Affairs. 2018 May 1;37(5):757-64.
13. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Medical preparedness for bioterrorism and chemical warfare: A public health integration review. Medicine. 2025 May 2;104(18):e42289.
14. McClellan KA, Avard D, Simard J, Knoppers BM. Personalized medicine and access to health care: potential for inequitable access?. European Journal of Human Genetics. 2013 Feb;21(2):143-7.
15. Edyedu I, Ugwu OP, Ugwu CN, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI. The role of pharmacological interventions in managing urological complications during pregnancy and childbirth: A review. Medicine. 2025 Feb 14;104(7):e41381.
16. Bosschieter TM, Xu Z, Lan H, Lengerich BJ, Nori H, Painter I, Souter V, Caruana R. Interpretable predictive models to understand risk factors for maternal and fetal outcomes. Journal of Healthcare Informatics Research. 2024 Mar;8(1):65-87.
17. Paul-Chima UO, Ugwu CN, Alum EU. Integrated approaches in nutraceutical delivery systems: optimizing ADME dynamics for enhanced therapeutic potency and clinical impact. RPS Pharmacy and Pharmacology Reports. 2024 Oct;3(4):rqae024.
18. Du Y, Rafferty AR, McAuliffe FM, Wei L, Mooney C. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Scientific Reports. 2022 Jan 21;12(1):1170.

CITE AS: Nakawungu Catherine (2026). Integrating Metabolomics with Family History for Preeclampsia Risk Prediction: Interpretability, Bias, and Real-World Performance, Implementation, and Equity Considerations. RESEARCH INVENTION JOURNAL OF SCIENTIFIC AND EXPERIMENTAL SCIENCES 6(1):35-43. https://doi.org/10.59298/RIJSES/2026/613543