

REVIEW OPEN ACCESS

Clinical Management

Prediction Models for Maternal and Offspring Short- and Long-Term Outcomes Following Gestational Diabetes: A Systematic Review

Yasmina Al Ghadban¹  | Nerys M. Astbury^{2,3}  | Abdallah Kurdi⁴  | Ankita Sharma¹ | Beatrice Ope^{5,6} | Tzu-Ying Liu⁷ | Lucy MacKillop^{1,8} | Huiqi Y. Lu⁹ | Jane E. Hirst^{1,6,8}

¹Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK | ²Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK | ³NIHR Oxford Biomedical Research Centre, Oxford University Hospitals, NHS Foundation Trust, Oxford, UK | ⁴Department of Biochemistry and Molecular Genetics, Faculty of Medicine, American University of Beirut, Beirut, Lebanon | ⁵Faculty of Medicine, School of Public Health, Imperial College London, London, UK | ⁶The George Institute for Global Health, Imperial College London, London, UK | ⁷Nuffield Department of Population Health, University of Oxford, Oxford, UK | ⁸Oxford University Hospitals, NHS Foundation Trust, Oxford, UK | ⁹Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Correspondence: Yasmina Al Ghadban (yasmina.alghadban@wrh.ox.ac.uk)

Received: 9 April 2024 | **Revised:** 4 November 2024 | **Accepted:** 26 March 2025

Funding: This study was supported by the MRC iCase Doctoral Training Program, the UKRI Future Leaders Fellowship; the Royal Academy of Engineering Daphne Jackson Research Fellowship (EP/N020774/1); the NIHR School of Primary Care Research, Diabetes UK (21/0006324); and the Diabetes Research and Wellness Foundation and NIHR Biomedical Research Centre Oxford.

Keywords: gestational diabetes | prediction models | prognosis modeling

ABSTRACT

Objectives: Gestational diabetes mellitus (GDM), affecting one in seven pregnant women worldwide, can have short- and long-term adverse outcomes for both the mother and her baby. Despite a raft of prognostic models aiming to predict adverse GDM outcomes, very few have impacted clinical practice. This systematic review summarizes and critically evaluates prediction models for GDM outcomes, to identify promising models for further evaluation.

Methods: We searched EMBASE, MEDLINE, Web of Science, CINAHL, and CENTRAL for studies that reported the development or validation of predictive models for GDM outcomes in mother or offspring (PROSPERO: CRD42023396697).

Results: Sixty-four articles detailing 103 developed and 12 validated models were included in this review. Of these, 45% predicted long term, 31% birth, and 23% pregnancy outcomes. Most models (87%) had a high risk of bias, lacking sufficient outcome events, internal validation, or proper calibration. Only eight models were found at low risk of bias.

Conclusions: Our findings highlight a gap in rigorously developed prediction models for adverse GDM outcomes. There is a need to further validate existing models and evaluate their clinical utility to generate risk prediction tools capable of improving clinical decision-making for women with GDM and their children.

1 | Introduction

Gestational diabetes mellitus (GDM), glucose intolerance with first onset or recognition during pregnancy, is one of the most common pregnancy complications affecting up to one in seven

pregnant women worldwide [1]. GDM can have short-term adverse consequences for the mother and the baby, such as increased risk of preterm delivery, macrosomia, and admission to neonatal intensive care unit (NICU) [2]. Additionally, although most women diagnosed with GDM return to normoglycemia shortly after giving

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Obesity Reviews* published by John Wiley & Sons Ltd on behalf of World Obesity Federation.

birth, the condition can have long-term consequences for both the mother and the child. Women with GDM face an increased risk of developing Type 2 diabetes (T2DM): A meta-analysis of 20 studies found that women with GDM were 7.4 times as likely to develop T2DM than women with normoglycemic pregnancies [3]. Women with GDM are also at a significantly higher risk of hypertension, obesity, and cardiovascular morbidity [4, 5]. In addition, their children are at a heightened risk of obesity, cardiovascular morbidity, and glucose intolerance later in life [6–8].

There are interventions that hold the potential to mitigate many GDM-related complications, such as diet advice, lifestyle modification, and medications [9]. However, proper detection, treatment, and follow-up programs are resource-intensive and often face challenges with poor compliance. With increasing obesity and maternal age, GDM prevalence is likely to increase and continue to pose substantial burden on healthcare systems, necessitating the development of effective strategies for risk stratification, resource allocation, and personalized management [10].

In recent years, researchers have increasingly turned to data-driven prediction models combining various factors to estimate an individual's likelihood of developing a specific outcome. Prognostic models can be used to predict the risk of GDM preconception or in the early stages of pregnancy. Two recent reviews have studied and compared prognostic models predicting the risk of developing GDM [11, 12] and found that although several published models achieved acceptable discrimination and calibration, they varied in quality and lacked independent validation. After a diagnosis of GDM, prognostic models can also be used to predict the risk of GDM outcomes during pregnancy; at birth and in the postnatal period; and long term (Figure 1). Despite the growing interest and potential of prediction models for women with GDM, none to our knowledge is currently used widely in

clinical practice. A systematic review of five prediction models for adverse outcomes during pregnancy in women with GDM pregnancy found that all models were at high risk of bias in their development, with a lack of external validation limiting their clinical utility [13]. However, given the fast development in this field, there has been further work because this review was conducted. Additionally, one of the most important consequences of GDM is the association with longer term cardiometabolic complications. Although several prediction models have been published in this area, to our knowledge, no review has critically appraised the quality and clinical utility of this evidence.

As such, the aim of this systematic review was to critically evaluate published prognostic models for the prediction of short- and long-term outcomes associated with GDM, considering outcomes for both the woman and offspring. This synthesis will inform whether further testing of existing models is justifiable or the development of new models is needed before predictive models will make a measurable impact on clinical outcomes.

2 | Methods

The systematic review was registered on PROSPERO, the International Prospective Register of Systematic Reviews (CRD42023396697), and follows the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [14].

2.1 | Search Strategy

We searched EMBASE, MEDLINE, CENTRAL, Web of Science (core collection), and CINAHL. The search filters included keywords related to pregnancy, diabetes, and

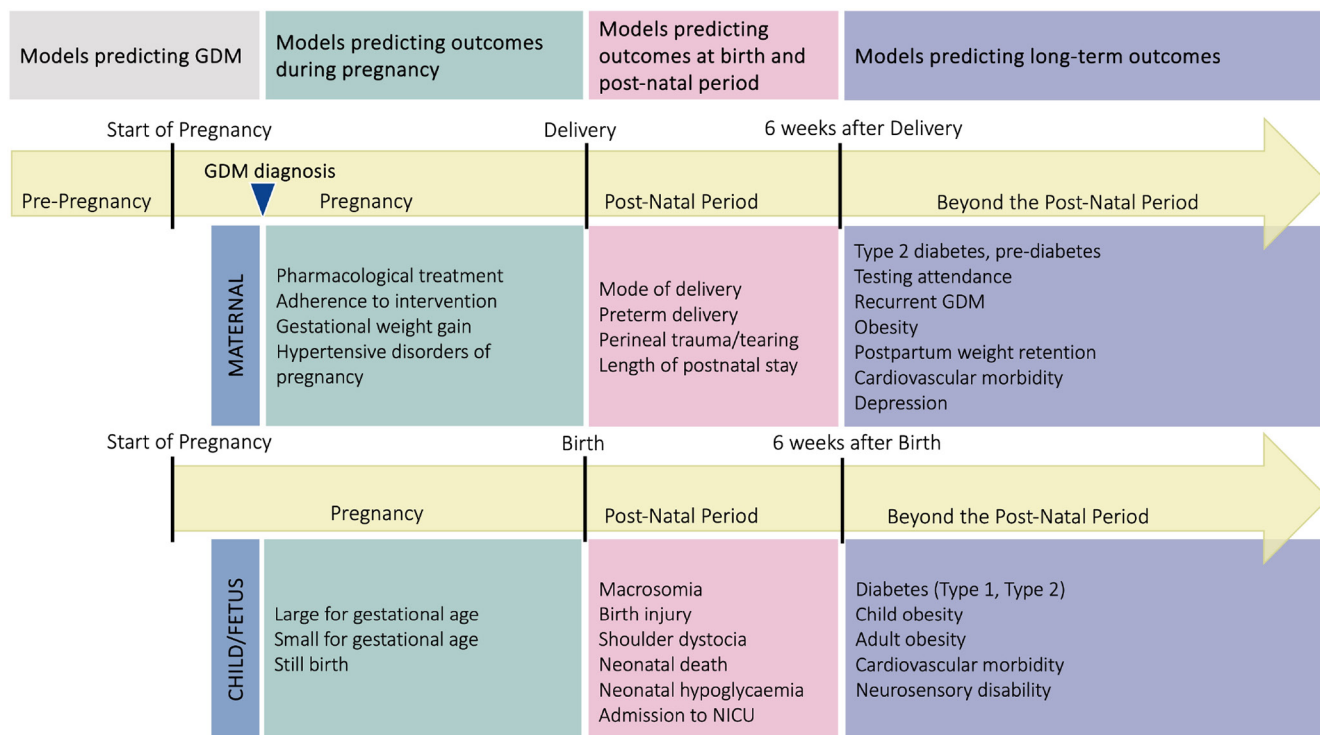


FIGURE 1 | The use of prediction models at different stages of pregnancy: preconception, during pregnancy, at birth and in the postnatal period, and long-term.

prognostic models and were developed based on previously validated search filters [15, 16]. No restrictions based on language were included in the strategy. Forward and backward citation searching was also performed for the studies included to improve the yield of the search strategies. An example search strategy is presented in Table S1. All databases were last searched on September 8, 2023.

2.2 | Selection Criteria

The eligibility criteria were formulated with the PICOTS (Population, Intervention, Comparator, Outcomes, Timing, Setting) framework provided by the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) (Table 1) [17].

TABLE 1 | Framing the review eligibility criteria with PICOTS.

PICOTS	
Type of model	1. Prognosis model studies without external validation in independent dataset 2. Prognosis model studies with external validation in independent dataset 3. External model validation studies with or without model updating
Population	Women with GDM as defined in the included studies and children born to mothers with GDM
Intervention	Prognostic model developed to predict maternal and child GDM outcomes
Comparator	Not applicable
Outcome	Maternal and child outcomes, listed in Figure 1, were selected based on outcomes studied in the literature [18] and a core outcome set developed for GDM [19]. Due to variations in outcome definitions, outcomes were collected as reported and defined in the included studies.
Timing	Models predicting pregnancy or birth outcomes will evaluate endpoints during the period between diagnosis with GDM until the end of the postnatal period (6 weeks after delivery). Models predicting long-term outcomes will evaluate endpoints at defined months or years after delivery.
Setting	The intended use of prognostic models is to perform risk stratification in the assessment of pregnant women to inform healthcare professionals' decision making and resource allocation. As such, the models are designed for use in clinical settings.

Model development studies with or without validation and external model validation studies with or without model updating were included if they were predicting at least one outcome of interest, as a single endpoint or composite, as defined in the specific studies (Figure 1).

We excluded studies reporting models with a single predictor, studies investigating overall prognosis or prognostic factors without model development or validation, studies that included only pregestational diabetes or only twin pregnancies; and non-original and non-peer-reviewed studies. Studies with both GDM and pregestational diabetes women were included only if the models/outcomes are reported separately for women with GDM. There were no restrictions on the type of data used, study design, timing, or setting of data collection.

2.3 | Selection Process

The results of the literature search were imported into Covidence, a systematic review management software, for deduplication and screening [20]. Title/abstract screening and full text screening were performed by two independent reviewers (YAG and AS), with a third reviewer (JEH) referred to for conflicts.

2.4 | Data extraction and Critical Appraisal

Data extraction was informed by CHARMS [17], and critical appraisal and assessment of applicability were done using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [21]. A template for data extraction and risk of bias assessment in systematic reviews of predictive models informed by CHARMS and PROBAST was used [22].

Specifically, for model development, we extracted the following items: study design; outcome definition and measurement; candidate predictors definition, measurement, and method of selection; included predictors; sample size; number of events; handling and reporting of missing data; type of model; and methods and metrics used for model evaluation. For studies reporting validation, we extracted similar items but also extracted differences between the model development and validation regarding definition and measurement of predictors and outcomes, handling of continuous predictors and methods used to handle missing data.

Data extraction and critical appraisal were performed once for each development and validation of each distinct prediction model in a publication. In a deviation to the published protocol, data extraction and critical appraisal were completed by the lead reviewer (YAG) and verified by a second reviewer (AK, BO, and TYL) for all models. Uncertainties were discussed between reviewers and referred to a third (JEH) when an agreement could not be reached.

2.5 | Data Synthesis

A meta-analysis was not applicable given the heterogeneity of studies included and the lack of external validation studies available for any one model. Instead, data were synthesized

qualitatively as reported in the protocol, using generated tables from the template of CHARMS and PROBAST [22].

3 | Results

3.1 | Study Selection

Our searches identified 11,543 records. After deduplication, 5964 studies were screened by title and abstract, and 137 studies were screened by full text screening. Sixty-four studies, describing 115 models (103 model development and 12 external validation), were included in this review (Figure 2).

3.2 | Included Study Characteristics

Of the 64 included studies, most were retrospective cohort studies (39/64; 61%), 16 (25%) were prospective cohort studies, 4 (6%)

were nested case-control studies, and 2 (3%) were randomized trials. Forty-one (64%) studies were single center studies.

Studies were from 24 different countries, with most represented countries (17/24, 71%) being high-income countries according to the World Bank 2023 classification [23]. Nearly half of the studies (30/64; 47%) were based in Europe and North America and the Caribbean. A significant proportion of studies (25/64; 39%) originated from the Western Pacific region, whereas fewer than 10% of studies represented South America, Africa, the MENA region (Middle East and North Africa), or Southeast Asia. Diagnostic criteria for GDM varied across studies and regions, with the most common being the Carpenter & Coustan criteria [24] and the International Association of Diabetes in Pregnancy Study Group (IADPSG) criteria [25].

The overwhelming majority of studies (56/64; 88%) were model development studies without external validation. Seven studies (11%) reported both model development and external validation,

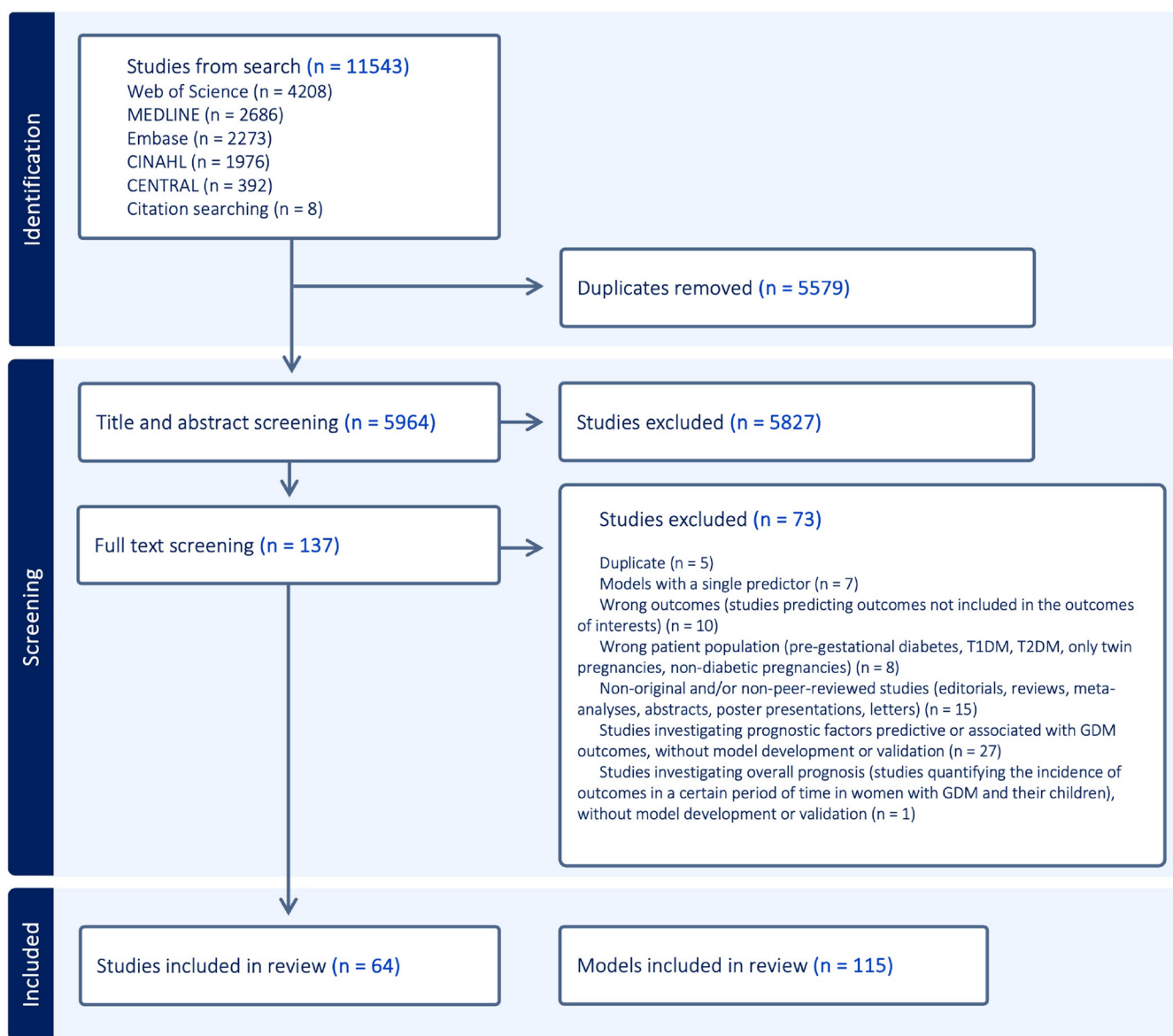


FIGURE 2 | PRISMA flow diagram of study selection.

and only one (2%) study described an external validation study without model development. The validation only article included several authors who were also on the paper that developed the model being evaluated. Thirty-nine (61%) studies reported one model, whereas the remaining 25 (39%) studies reported multiple models. Studies reported more than one model for a range of reasons: external validation, use of different predictors or populations, prediction of different outcomes, or time of outcome ascertainment.

Long-term outcomes were most common, representing just under half of all studies (29/64; 45%). Among maternal cardiometabolic complications, outcomes predicted included Type 2 diabetes in 19 studies (30%), glucose metabolism complications (postpartum glucose intolerance, prediabetes, and abnormal glucose tolerance) in 8 studies (13%), and lipid metabolism issues in 1 study (2%). Long-term outcome studies exclusively focused on maternal health with no model predicting long-term offspring outcome. Birth outcomes were predicted in 20 studies (31%) with macrosomia predicted in 10 studies. Eleven studies specifically targeted offspring birth outcomes and an additional five studies examining both maternal and offspring outcomes. Finally, pregnancy outcomes were predicted in 15 studies (23%) with 14 studies predicting treatment modality. Almost all studies reported single outcomes, with only three studies predicting composite outcomes.

Reporting of participant characteristics varied by study. Around half of the studies reported the number of nulliparous women (35/64; 55%) and the population's ethnic composition (32/64; 50%). Additionally, excluding studies that aimed to predict GDM treatment modality, 29/47 studies (62%) reported GDM treatment. One study included only women on glyburide [26], three studies did not include women receiving any pharmacological treatment [27, 28, 29], and one study women received no treatment for GDM (including medical nutrition therapy) as both clinicians and patients were blinded to the oral glucose tolerance test (OGTT) results [30].

The full details of included studies are available in Table 2.

3.3 | Sample Size

In 93 of 103 developed models (90%), data on the number of events and candidate predictors allowed EPV calculations, which ranged from 0.7 to 151.3 with a median of 5.5. Fewer than half (45/93; 48%) of these models met an $EPV \geq 10$, and only a third (32/93; 34%) met an $EPV \geq 20$. For the 12 validated models, four included fewer than 100 participants with the outcome.

3.4 | Missing Data

An overwhelming majority did not report the number of missing data (93/115; 81%) or the handling of missing data (49/115; 43%). Of the models that reported how missing data were handled, over half (39/66; 59%) used complete-case analysis, 18/66 (27%) used multiple imputation, and 5/66 (8%) used single imputation.

3.5 | Model Development

The included studies described 103 developed models with varying model development methods (Table S2). The sample size used to develop the prediction models ranged from 40 to 90,143 with a median of 322, and the number of events ranged from 4 to 10,289 with a median of 78. Most models used logistic regression (74/103; 72%) and 15 (15/103, 15%) used machine-learning methods. Among the 15 machine learning models, tree-based models were the most common, with six models employing techniques such as decision trees, CART (Classification and Regression Trees), random forest, and RECPAM (Recursive Partition and AMalgamation). One study used a probabilistic model, specifically the Naive Bayes classifier. Two models were ensemble models, utilizing a voting meta-algorithm to combine top-performing classifiers. Neural networks were used in two models, applying deep learning techniques to capture complex relationships in GDM-related data. Lastly, specialized methods were used in four studies, including cluster analysis, a filtered classifier, AIRS (Artificial Immune Recognition System), and OPLS-DA (Orthogonal Partial Least Squares Discriminant Analysis) for classification tasks. From the 45 models predicting outcomes determined after months or years of follow up, only 7/45 (16%) models used Cox regression, a survival analysis method that can account for loss to follow up. In the 84 models that used regression-based methods, only nine models (11%) applied shrinkage techniques to alleviate overfitting. Additionally, less than half (35/84; 42%) reported regression coefficients, and only 20/84 (24%) reported the full model equation including the intercept or baseline survival.

3.6 | Predictors

In model development, 47/103 (46%) models selected candidate predictors based on univariable associations, whereas 28/103 (27%) selected candidate predictors based on existing knowledge.

The number of predictors in the model ranged from 2 to 20, with a median of 5. A variety of candidate predictors were selected across models (Table S2). Clinical factors routinely available from standard electronic health records (BMI, age, ethnicity, parity, and medical history) and blood glucose measures recorded at or around the time of GDM diagnosis (Hb1Ac, OGTT results) were the most used. However, 29 (28.2%) models also used predictors not typically available in clinical practice, such as omics (proteomics, metabolomics, and transcriptomics), genetic factors, or digital daily blood glucose readings. The single most commonly used predictor was BMI, included in 70 models (68%), followed by maternal age included in 49 models (48%), and fasting plasma glucose included in 47 models (46%) (Figure 3).

Studies typically reported the mean, standard deviation, or interquartile range of continuous predictors. Only 10 models (10/103; 10%) reported the range of any continuous predictors. Additionally, only half of models (51/103; 50%) handled continuous variables without categorization. In 21/27 (78%) of models that reported categorizing continuous variables, cut points were chosen by selecting the thresholds that result in the highest predictive performance, rather than by predefining thresholds.

TABLE 2 | Characteristics of the studies included in the systematic review ($n = 64$).

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Pregnancy outcomes									
Phaloprakarn, 2009 [31]	Retrospective cohort ($n = 813$)	Jan 2003 to Feb 2008	Single center; Bangkok, Thailand	Universal screening; Carpenter & Coustan	92% Thai; 8% Southeast Asians	40.0%	34.1%	1	Pre-eclampsia
Badri, 2022 [32]	Retrospective cohort ($n = 1298$)	Jan 2018 to Dec 2018	Single center; Brussels, Belgium	IADPSG	85.0% Europe, North Africa, Middle East; 12.0% Sub-Saharan Africa; 2.9% South or East Asia	27.3%	NR	4 (2 different screening used and external validation)	Treatment modality
Barnes, 2016 [33]	Retrospective cohort ($n = 4015$)	1992 to 2015	Single center; New South Wales, Australia	Universal screening; ADIPS 1998	33.7% East and South-East Asian; 27.0% Middle Eastern; 22.5% European; 11.5% South Asian; 5.3% African or Pacific Islander	33.0%	32.4%	2 (external validation)	Treatment modality
Du, 2021 [34]	Retrospective cohort ($n = 626$)	May 2016 to Dec 2019	Single center; Sheng Jing Hospital, China	IADPSG	100% Chinese	81.5%	30.0%	1	Treatment modality
Eleftheriades, 2021 [35]	Prospective cohort ($n = 775$)	2010 to 2018	Single center; Athens, Greece	Universal screening; IADPSG	NR	NR	16.8%	2 (external validation)	Treatment modality
Ford, 2022 [36]	Retrospective cohort ($n = 2048$)	Jan 2016 to Dec 2017	Single center; Melbourne, Australia	Universal screening; IADPSG	NR	29.5%	31.6%	1	Treatment modality
Harper, 2016 [26]	Retrospective cohort ($n = 220$)	Jan 2007 to Dec 2013	Single center; Alabama, USA	Universal screening; Carpenter & Coustan	51.4% Black; 12.3% White; 28.6% Hispanic; 41.8% Other	27.7%	100.0%	1	Treatment modality

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Kofoed, 2023 [37]	Retrospective cohort (n = 1104)	2012 to 2017	Single center; Denmark	Selective screening; Danish Guidelines	86.9% Caucasian; 12.1% Afro-Caribbean; 1.1% Other	44.4%	25.5%	1	Treatment modality
Liao, 2022 [38]	Retrospective cohort (n = 27,240)	Jan 2007 to Dec 2016	Multicenter; California, USA	Universal screening; Carpenter & Coustan	22.7% White; 28.1% Hispanic; 4.3% African American; 40.3% Asian/Pacific Islander; 4.6% Other	40.8%	37.8%	8 (different predictors and external validation)	Treatment modality
Mendez-Figueroa, 2014 [39]	Retrospective cohort (n = 367)	Jun 2006 to Dec 2011	Single center; Rhode Island, USA	Carpenter & Coustan	23.5% White; 13.7% African American; 15.0% Asian; 39.0% Hispanic; 8.7% Other	18.3%	38.9%	1	Treatment modality
Much, 2015 [40]	Retrospective cohort (n = 856)	1998 to 2010	Single center; Munich, Germany	IADPSG	NR	NR	21.4%	1	Treatment modality
Souza, 2018 [41]	Retrospective cohort (n = 408)	2012 to 2015	Single center; Sao Paulo, Brazil	IADPSG	NR	NR	33.1%	1	Treatment modality
Velardo, 2021 [42]	Retrospective cohort (1543)	2014 to 2019	Multicenter; NHS Trusts, UK	Varies by center	NR	NR	NR	1	Treatment modality
Yerlikaya, 2018 [43]	Retrospective cohort (n = 203)	May 2015 to Jan 2017	Single center; Vienna, Austria	Universal screening; IADPSG	NR	27.6%	46.8%	3 (different predictors)	Treatment modality
Zaccara, 2023 [44]	Retrospective cohort (n = 869)	Jan 2012 to Mar 2020	Single center; Sao Paulo, Brazil	Universal screening; IADPSG	NR	NR	16.7%	1	Treatment modality
Birth outcomes									
Pintaudi, 2018 [45]	Retrospective cohort (n = 2736)	Jan 2012 to May 2015	Multicenter; Italy	Selective screening; IADPSG	Caucasian 44.8%	45.3%	41.3%	1	Adverse outcomes composite

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Cooray, 2022 [46]	Retrospective cohort (<i>n</i> = 1747)	Jul 2017 to Jun 2018	Multicenter; Melbourne, Australia	Universal screening; IADPSG	19.6% Caucasian; 35.3% Southern and Central Asian; 13.8% East Asian; 5.2% African; 3.3% Oceanian not Australian; Other 22.9%	38.2%	36.6%	2 (external validation)	Adverse outcomes composite
Park, 2015 [47]	Retrospective cohort (<i>n</i> = 802)	Mar 2001 to Apr 2013	Single center; Seoul, South Korea	Carpenter & Coustan	NR	NR	NR	1	Adverse outcomes composite
Lu, 2022 [48]	Randomized trial (<i>n</i> = 155)	Sept 2013 to Jun 2015	Single center; Oxford, UK	IADPSG	NR	NR	NR	2 (different predictors)	Cesarean delivery
Phaloprakarn, 2020 [31]	Retrospective cohort (<i>n</i> = 385)	Jan 2011 to Dec 2014	Single center; Bangkok, Thailand	Universal screening; Carpenter & Coustan	NR	41.6%	12.2%	2 (external validation)	Cesarean delivery
Ramos, 2023 [49]	Retrospective cohort (<i>n</i> = 3570)	Jan 2022 to Mar 2013	Single center; USA	Carpenter & Coustan	62.4% White; 7.2% Black; 5.9% Asian; 24.5% Other	55.2%	24.9%	1	Cesarean delivery
Kim, 2022 [50]	Retrospective cohort (<i>n</i> = 660)	2006 to 2013	Multicenter; South Korea	Carpenter & Coustan	NR	NR	NR	1	Large for gestational age
Kang, 2019 [51]	Retrospective cohort (<i>n</i> = 1891)	Jan 2010 to Dec 2017	Single center; Nantong, China	NR	NR	NR	NR	2 (different predictors)	Macrosomia
Odinokova, 2019 [52]	Prospective cohort (<i>n</i> = 220)	2017 to 2018	Moscow, Russia	IADPSG	100% Caucasian	NR	NR	1	Macrosomia
Sun, 2020 [53]	Prospective cohort (<i>n</i> = 64)	Jun 2017 to Apr 2018	Single center; Hangzhou, China	IADPSG	100% Chinese Han	76.4%	NR	1	Macrosomia
Sun, 2023 [54]	Retrospective cohort (<i>n</i> = 991)	Nov 2020 to Feb 2022	Single center; Shanxi, China	IADPSG	NR	NR	9.3%	1	Macrosomia

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Tomlinson, 2018 [55]	Retrospective cohort (n = 275)	Mar 2010 to May 2012	Massachusetts, USA	Carpenter & Coustan	82.0% White; 5.0% Black; 12.0% Other	51.0%	NR	1	Macrosomia
Tomlinson, 2021 [56]	Retrospective cohort (n = 477)	Jan 2015 to Jun 2018	Perinatal database, St Louis, USA	Carpenter & Coustan	54.0% White; 40.0% Black; 6.0% Other	30.0%	NR	1	Macrosomia
Yuan, 2022 [57]	Unclear (n = 40)	Jul 2019 to Jun 2021	Multicenter; Jiangsu, China	IADPSG	NR	NR	NR	1	Macrosomia
Yuan, 2023 [27]	Retrospective cohort (n = 88)	Jan 2020 to Jun 2020	Single center; Nanjing, China	IADPSG	100% Han Chinese	70.5%	0%	1	Macrosomia
Zhang, 2023 [28]	Retrospective cohort (n = 322)	Oct 2020 to Oct 2021	Single center; Shantou City, China	Universal screening; IADPSG	NR	NR	0%	1	Macrosomia
Zou, 2021 [58]	Retrospective cohort (n = 783)	Oct 2019 to Oct 2020	Single center; Qingdao, China	IADPSG	NR	NR	NR	1	Macrosomia
Huang, 2023 [59]	Retrospective cohort (n = 564)	Jan 2017 to Jan 2020	Single center; Wenzhou, China	IADPSG	NR	63.5%	10.5%	2 (external validation)	Preterm delivery
Mukherjee, 2023 [60]	Retrospective cohort (n = 627)	2015 to 2016	UK	Selective screening; IADPSG	NR	NR	33.2%	7 (different outcomes)	Preterm birth, hypoglycemia, hyperbilirubinemia, NICU, macrosomia, shoulder dystocia, composite
McIntyre, 2018 [30]	Prospective cohort (n = 1248)	Jul 2000 to Apr 2006	Single center; South Brisbane, Australia	ADIPS 1991; IADPSG; Carpenter & Coustan; NICE 2015	100% Caucasian	54.5%	NR	6 (different outcomes)	Cesarean delivery, birth injury, LGA, adiposity, hyperinsulinemia, hypoglycemia
Long-term outcomes									
Jotic, 2023 [29]	Prospective cohort (n = 147)	Jan 2017 to Dec 2021	Single center; Serbia	IADPSG	NR	NR	0%	1	Dyslipidemia

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Hahn, 2023 [61]	Retrospective cohort (n = 159)	Jan 2014 to Sept 2020	Single center; Germany	IADPSG	NR	NR	59.7%	1	GDM recurrence
Schwartz, 2016 [62]	Retrospective cohort (n = 788)	1991 to 2012	Single center; Israel	Universal screening; Carpenter & Coustan or NDDG	52.8% Jews; 47.2% Arabs	48.5%	34.1%	1	GDM recurrence
Bengtson, 2022 [63]	Prospective cohort (n = 203)	Jan 2017 to Jul 2018	Single center; Rhode Island, USA	Carpenter & Coustan	48.3% White; 10.3% Non-Hispanic black; 30.5% Hispanic; 10.8% Other	NR	57.1%	1	Postpartum glucose intolerance
Kondo, 2018 [64]	Retrospective cohort (n = 123)	Apr 2012 to Apr 2017	Single center; Yamaguchi-City, Japan	Universal screening; IADPSG	NR	NR	22.0%	3 (different predictors)	Postpartum glucose intolerance
Bartakova, 2021 [65]	Retrospective cohort (n = 244)	2011 to 2013	Single center; Czech Republic	Unclear	100% Caucasian	44.3%	37.3%	1	Postpartum glucose intolerance
Pei, 2023 [66]	Prospective cohort (n = 133)	Jan 2018 to Dec 2020	NR	IADPSG	NR	NR	NR	1	Impaired glucose tolerance remission
Muche, 2020 [67]	Prospective cohort (n = 112)	Mar 2018 to Mar 2019	Multicenter; Gonda, Ethiopia	Universal screening; IADPSG	NR	NR	6.2%	1	Postpartum prediabetes
Periyathambi, 2022 [68]	Retrospective cohort (n = 607)	Jan 2016 to Dec 2019	Single center; Nuneaton, UK	Selective screening; NICE 2015	79.2% White European; South Asian 11.1%; Other 9.7%	43.9%	NR	1	Return for testing
Ingram, 2017 [69]	Retrospective cohort (n = 165)	Jan 2007 to Jun 2009	Single center; Tasmania	Universal screening; ADIPS 1998	86.6% Caucasian; 4.8% Aboriginal and Torres Strait; 8.5% Other	NR	60.6%	3 (different outcome)	Return for testing, abnormal glucose tolerance

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Barden, 2013 [70]	Prospective cohort (n = 150)	1998 to 2001	Multicenter; Perth, Western Australia	ADIPS 1998	74.7% Caucasian; 25.3% Asian/aboriginal/middle-eastern	18.7%	NR	1	Type 2 diabetes
Lin, 2011 [71]	Retrospective cohort (n = 607)	NR	Single center; Taipei, Taiwan	Universal screening; NDDG	NR	NR	NR	1	Type 2 diabetes
Man, 2021 [72]	Randomized trial (n = 317)	1996 to 2001	Multicenter; USA	NR	59.3% Caucasian; 19.9% African American; 17.0% Hispanic; 3.8% Other	NR	33.1%	1	Type 2 diabetes
Ignell, 2016 [73]	Prospective cohort (n = 362)	2003 to 2005	Multicenter; Skane, Southern Sweden	Universal screening; Modification to WHO 1999	16.0% non-European	NR	9.1%	2 (different population)	Type 2 diabetes
Allalou, 2016 [74]	Nested case-control (n = 244)	Sep 2008 to Dec 2011	Multicenter; California, USA	Carpenter & Coustan	17.2% White; 30.7% Asian; 11.1% Black; 39.3% Hispanic; 1.6% Other	35.2%	46.3%	3 (different predictors)	Type 2 diabetes
Joglekar, 2021 [75]	Prospective cohort (n = 82)	Jun 2003 to Dec 2005	Single center; Australia	ADIPS 1998	NR	NR	48.1%	2 (different predictors)	Type 2 diabetes
Khan, 2019 [76]	Nested case-control (n = 140)	Sept 2008 to Dec 2011	Multicenter; California, USA	Carpenter & Coustan	64.3% Hispanic; 35.7% Asian	23.6%	42.2%	1	Type 2 diabetes
Kwak, 2013 [77]	Prospective cohort (n = 395)	Jan 1996 to Feb 2003	Single center; Seoul, South Korea	Universal screening; NDDG	NR	NR	28.4%	4 (different predictors)	Type 2 diabetes
Lai, 2020 [78]	Nested case-control (n = 337)	Sept 2008 to Dec 2011	Multicenter; California, USA	Carpenter & Coustan	17.9% White; 29.8% Asian; 11.7% Black; 39.4% Hispanic; 1.2% Other	35.2%	46.3%	1	Type 2 diabetes
Lappas, 2015 [79]	Prospective cohort (n = 104)	Jun 2003 to Dec 2005	Single center; Australia	ADIPS 1998	NR	NR	48.1%	3 (different predictors)	Type 2 diabetes

(Continues)

TABLE 2 | (Continued)

Author, year	Source of data (sample size)	Enrolment period	Study setting	GDM diagnosis criteria	Participant characteristic ^a			Number of models (reasons for multiple models)	Outcome (s)
					Ethnicity	Nulliparous	Pharmacological treatment		
Lappas, 2016 [80]	Prospective cohort (n = 95)	Jun 2003 to Dec 2005	Single center; Australia	ADIPS 1998	NR	NR	51.0%	3 (different predictors)	Type 2 diabetes
Lappas, 2018 [81]	Prospective cohort (n = 98)	Jun 2003 to Dec 2005	Single center; Australia	ADIPS 1998	NR	NR	49.5%	3 (different predictors)	Type 2 diabetes
Zhang, 2021 [82]	Nested case-control (n = 216)	Sept 2008 to Dec 2011	Multicenter; California, USA	Carpenter & Coustan	16.2% White; 31.5% Asian; 8.3% Black; 43.1% Hispanic; 0.9% Other	NR	42.6%	1	Type 2 diabetes
Li, 2018 [83]	Retrospective cohort (n = 1263)	2005 to 2009	Single center; Tianjin, China	Universal screening; WHO 1999	100% Chinese	NR	NR	2 (different time for outcomes)	Type 2 diabetes
Elhour, 2006 [84]	Unclear (n = 165)	Dec 2001 to Jun 2002	Single center; UAE	Universal screening; Carpenter & Coustan	NR	NR	NR	1	Type 2 diabetes
Kohler, 2016 [85]	Prospective cohort (n = 304)	1989 to 1999	Multicenter; Germany	German Diabetes Association	NR	NR	33.5%	1	Type 2 diabetes
Marozas, 2018 [86]	NR (n = 151)	NR	NR	NR	NR	NR	14.4%	2 (different outcomes)	Type 2 diabetes and impaired metabolism
Cormier, 2014 [87]	Retrospective cohort (n = 214)	Oct 2009 to Aug 2012	Quebec City, Canada	NR	93.6% White	NR	NR	2 (different outcomes)	Type 2 diabetes and prediabetes
Ukah, 2022 [88]	Retrospective cohort (n = 90,143)	Apr 1989 to Mar 2016	Multicenter; Quebec, Canada	Varies by center	NR	60.3%	NR	2 (different time for outcomes)	Type 2 diabetes complications

^aWhen possible, in studies that reported patient characteristics for each subgroup, data were converted to get a value for the complete sample.

Abbreviations: ADIPS, Australasian Diabetes in Pregnancy Society; IADPSG, International Association of Diabetes in Pregnancy Study Groups; LGA, large for gestational age; NDDG, National Data Diabetes Group; NICE, National Institute for Health and Care Excellence; NICU, Neonatal Intensive Care Unit; NR, not reported; OGTT, Oral Glucose Tolerance Test; WHO, World Health Organization.

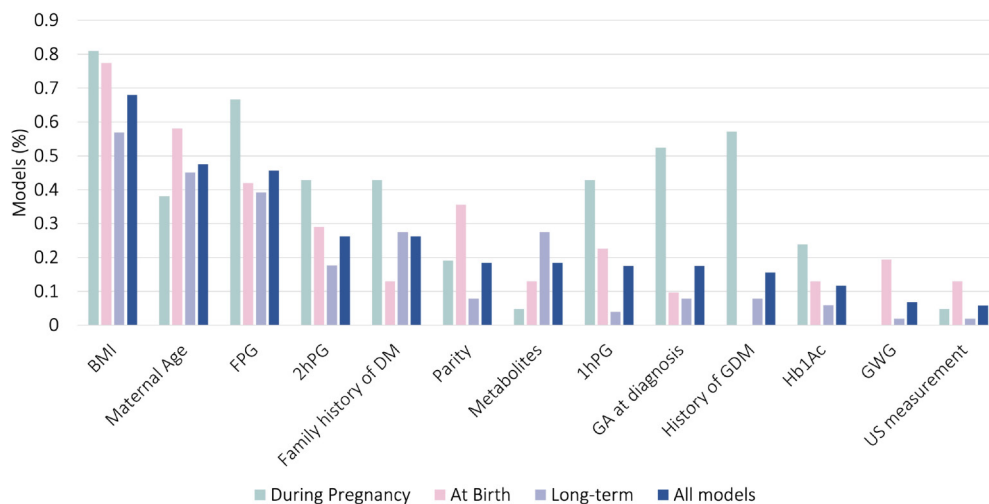


FIGURE 3 | Bar graph showing the most common predictors included in the development of the models ($n = 103$) by type of model.

3.7 | Performance Metrics

Calibration metrics were reported in 30 models (26%). Among these, nine models (30%) employed the Hosmer–Lemeshow goodness of fit test. Only one model reported the calibration slope, whereas 20 included a calibration plot. R^2 values were reported in three models. Discrimination, measured through the C-statistic (area under the ROC curve), was reported in 108 models (94%). Additionally, 60% of models (69/115) presented the ROC curve, but only 16% (11/69) labeled cut-off points to determine sensitivity and specificity, whereas 51% (59/115) reported these metrics. Clinical utility—assessing the actual usefulness in decision-making—was examined for 10/115 (9%) models using decision curve analysis (DCA), a novel method to calculate the net-benefit of a model in comparison to default strategies such as treating all or none.

3.8 | Model Presentation

Thirty-five models (35/103; 34%) presented an alternative model presentation. Seventeen models (16/103; 16%) were presented as a score system, a simplified format where predictor coefficients are converted to integers and summed together. Eleven models (11/103; 11%) presented their model using a nomogram, a graphical representation that displays the associations between predictors and response in a multivariable model and enable users to estimate the outcome probability for each patient. The remaining models were presented using an online calculator or regression trees. Only 5/35 (14%) models presented both the full model equation and the simplified model.

3.9 | Model Validation

Most models lacked internal or external validation. Of the 103 developed models, 43 (42%) included one form of internal validation. Nine of these used random sample splitting between development and validation, whereas 34 applied cross-validation or bootstrap methods. External validation was reported in 12 models (12%). Clinical utility was evaluated in

10 models (9%), with decision curve analysis (DCA) applied to assess net benefit.

Of the 12 externally validated models, seven had an alternative presentation, but it was unclear whether the full model or the alternative model was validated. Nine of these models (9/12; 75%) used temporal validation, evaluating their model in data from the same center but from a different time period. Only three models (25%) were validated in an independent sample, and none were validated by independent researchers. All studies reported the sample size used to validate the model (range: 168 to 3234; median: 716), but two studies did not report the number of events. For validation studies that reported the number of events, the median number of events was 368 (range 628–1528). For 10/12 models, the characteristics of both the development and validation cohorts were reported, allowing a comparison of case mix between the two.

3.10 | Risk of Bias and Applicability

Most models (100/115; 87%) had an overall high risk of bias (Figure 4A). Low risk of bias was observed across the outcome, predictors, and participants domains; however, 99 models exhibited high risk of bias in the analysis domain. Less than half of models (52/115; 45%) were applicable to the research question (Figure 4B). Table S3 includes individual PROBAST domain ratings across all development and external validation analyses.

3.11 | Gray Literature

We identified 13 conference abstracts relevant to this review [89–101]. The findings of six of these abstracts were subsequently published in peer-reviewed articles included in this review [33, 46, 49, 50, 72, 75]. The limited information available in the abstracts hindered our ability to conduct adequate data extraction and critical appraisal. However, with the information available, we found similar methodological issues as in the included models such as inadequate evaluation of calibration, lack of internal validation, and small sample sizes.

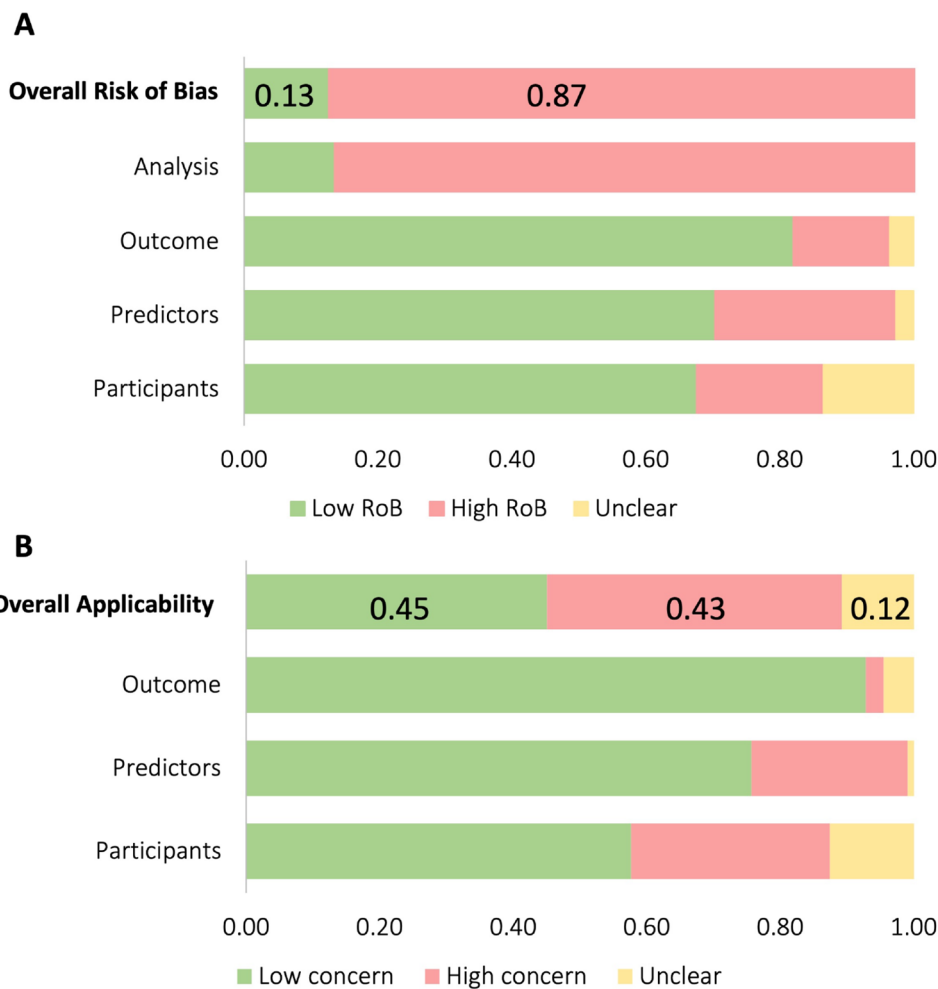


FIGURE 4 | Summary of risk of bias (A) and applicability (B) assessment.

4 | Discussion

4.1 | Summary of Findings

This systematic review is the largest review to date of prognostic models, which aim to determine the risk of adverse outcomes in women affected by GDM and her offspring. We included 64 studies, describing 115 models. The most common outcomes predicted were (i) the longer term risk of developing diabetes for the mother; (ii) birth outcomes, with a focus on macrosomia; and (iii) GDM treatment modality during pregnancy (diet or medication).

The regional distribution of studies revealed a concentration in Europe, North America, and the Western Pacific, where GDM prevalence is 7.2%, 7.8%, and 14.7% [1]. However, other regions with some of the highest reported GDM incidence rates—such as MENA (27.6%) and Southeast Asia (20.8%)—are under-represented in the literature, with fewer than 10% of studies originating from these areas [1]. This imbalance may limit the generalizability of prediction models and underscores the need for more research in regions with a high GDM burden to ensure broader applicability and relevance across diverse populations.

Additionally, a significant gap was identified for models addressing long-term health outcomes both for children born to mothers with GDM and for maternal cardiovascular risks.

Children exposed to GDM are at higher risk for obesity, metabolic syndrome, and Type 2 diabetes [6–8], yet few models account for these intergenerational risks, reducing opportunities for early intervention that could improve long-term health outcomes. Similarly, models that predict long-term cardiovascular outcomes for women with GDM are limited, despite the known association between GDM and increased risks of hypertension and cardiovascular disease [4, 5]. Developing models that incorporate these risks could support preventive strategies for both mother and child, ultimately enhancing lifetime care through targeted interventions and comprehensive management of long-term health risks associated with GDM.

Together, these gaps in regional representation, long-term offspring outcome modeling, and models of cardiovascular outcomes highlight important limitations in the current body of GDM prediction research. Addressing these areas in future studies will be essential to develop more comprehensive, generalizable, and clinically useful models that support individualized care across diverse settings and patient populations.

4.2 | Bias and Methodological Limitations

Overall, we found that most models are not suitable for clinical use in their current form due to major methodologic limitations.

Of the 115 models, eight unique models were found to be at low risk of bias. These models predicted treatment modality [38], a composite of adverse outcomes [46], Type 2 diabetes complications [88], and primary cesarean section [49, 102]. External validation in different populations is needed to determine real-world performance and impact.

These methodological challenges are consistent with findings in GDM research [13, 103] and other fields, such as of psychiatry, tuberculosis, and asthma [104–106], where prediction models frequently exhibit high risk of bias and inadequate calibration. The key methodological issues identified were as follows:

1. **Sample size:** Many models were developed with a low EPV, with fewer than half meeting the $EPV \geq 10$ criterion and even fewer reaching the recommended $EPV \geq 20$. These limitations raise concerns about overfitting, which can lead to overly optimistic model performance that fails to generalize to broader populations. To enhance the reliability of future models, studies should move beyond a one-size-fits-all EPV threshold. Instead, sample size requirements should be calculated based on the specific complexity of each model and the expected predictive accuracy [107, 108]. Adopting these tailored sample size calculations will help produce more robust and clinically applicable models across diverse patient populations.
2. **Handling of missing data:** The reporting and handling of missing data were frequently insufficient, with over half of the models relying on complete-case analysis. This approach can introduce significant bias by selectively excluding certain patients, resulting in a nonrandom subset of the study population, which can lead to skewed predictor-outcome associations and biased model performance. Multiple imputation is regarded as the most appropriate method for handling missing data in prediction model development and validation [109–111]. It minimizes bias by estimating the missing values based on other observed data, thereby preserving the overall structure of the dataset and providing more accurate estimates of model performance, including correct standard errors and *p* values. Future studies should prioritize multiple imputation or other rigorous approaches for handling missing data, as these methods provide better precision and less bias. Additionally, it is crucial for authors to transparently report how missing data were handled and to provide comparisons between included and excluded participants to assess potential bias.
3. **Predictor selection:** In our review, many models relied on univariable analysis to select candidate predictors, an approach that selects predictors based solely on their individual statistical significance without considering how they interact with other variables. This method can result in omitting relevant predictors that may only show importance when adjusted for other factors or, conversely, including predictors based on accidental associations. Such selection practices increase the risk of both underfitting and overfitting, compromising the model's performance in new datasets. A more reliable approach to predictor selection emphasizes the use of

nonstatistical methods grounded in established knowledge, clinical credibility, and practical considerations relevant to the model's intended setting. Predictors known to be significant in prior research should be retained, regardless of their statistical significance in a specific dataset, ensuring that the model incorporates clinically meaningful variables. Methods like backward selection or penalized regression can also help refine predictor sets in a multivariable context without relying on preliminary univariable testing, preserving predictor integrity while minimizing overfitting [112]. Additionally, using internal validation techniques to assess overfitting and optimize model performance is crucial when refining predictor sets [112, 113, 114]. Adopting these predictor selection strategies can improve model stability and enhance generalizability, ultimately producing models that are both simpler and more clinically relevant without compromising predictive power.

4. **Model evaluation:** Our review identified calibration as a frequently neglected aspect in model evaluation, with many studies relying solely on the Hosmer–Lemeshow test. This test is limited in statistical power, especially with large datasets, and can be overly sensitive to sample size, often leading to misleading assessments of calibration quality. Calibration plots and calibration slopes are more informative alternatives, as they visualize and quantify the alignment between predicted and observed outcomes across the risk spectrum, offering a more nuanced understanding of model performance [115, 116]. In addition, cut-off points on ROC curves were often overlooked. Without these measures, the interpretation of discrimination—the model's ability to differentiate between individuals with and without the outcome—remains incomplete. Reporting only the area under the ROC curve (AUC) can obscure important calibration issues, as a model with high discrimination may still miscalibrate, predicting risks inaccurately in certain subgroups. To improve future models, a more comprehensive approach to evaluating calibration and discrimination is needed. Calibration slopes, plots, and recalibration techniques should be used alongside discrimination metrics to provide a balanced assessment of a model's predictive accuracy.

We highlight two studies [46, 88] that used rigorous methods and transparent reporting in Boxes S1 and S2. External validation of these models in different populations is needed to determine real-world performance and impact.

4.3 | Implications for Research and Clinical Practice

The increasing availability of “big data” from EHR records, glucose sensors, and other sources (such as “omics studies”) has led to the development of a large number of prognostic models for GDM outcomes. Unfortunately, the lack of external validation of these models results in considerable research waste and limits the uptake of models in practice. Instead of continually developing new models, it is imperative to focus on the evaluation of

existing models in new datasets. Poor performance in external validation does not necessarily indicate the need for an entirely new model. Updating or recalibrating the existing model based on both development and validation data can improve its performance, stability, and generalizability.

When the development of a new model is warranted, rigorous methodology should be followed using existing guidelines for prognosis research. Methodological guidance is available for all aspects of model development including sample size calculation, selection of predictors, shrinkage techniques, model evaluation, and internal and external validation [1–3]. Models should be reported adequately using the TRIPOD guidelines to ensure transparency and reproducibility [117].

Finally, impact studies are critical to understanding a model's true clinical value. By evaluating how a model affects clinical decision-making, patient outcomes, resource utilization, and cost-effectiveness, impact studies provide insight into the practical benefits of these tools [118]. Such studies are crucial for determining whether a model delivers measurable improvements in care, guiding the integration of prognostic models into practice confidently. These approaches collectively support the translation of GDM prognostic models into tools that can reliably enhance patient care and clinical efficiency.

4.4 | Strengths and Limitations

This review demonstrates several notable strengths, including a prospective protocol registration, an extensive search across multiple databases without limitations on language or publication date and data extraction and critical appraisal using specialized tools such as CHARMS [17] and PROBAST [21]. Additionally, although nonexhaustive, the gray literature search increased the review's comprehensiveness and reduced potential publication bias. However, we were unable to assess for reporting bias as only one included study followed a prospectively published protocol. This further drives the point of the need for more transparent reporting in predictive modeling studies. An important limitation is that we were unable to compare prediction models quantitatively due to the heterogeneity of studies included. For example, GDM was diagnosed using eight different criteria and management protocols differed by center. Focusing on external validation can support future head-to-head comparisons of prognostic models.

5 | Conclusion

In conclusion, this systematic review provides a comprehensive summary of current prediction models for GDM outcomes and establishes a benchmark for future model development. By identifying key methodological best practices and emphasizing the need for improved geographic representation, this review highlights critical areas for enhancement in the field. Adopting these practices—such as rigorous validation, tailored sample size calculations, and broader regional applicability—will be essential for improving the reliability, generalizability, and clinical utility of future models, ultimately supporting more effective, patient-centered care in diverse global settings.

Author Contributions

YAG and JEH conceived of the study. YAG was the lead reviewer for the abstract/title screening, full text screening, data extraction, and bias assessment and led the writing of both the protocol and manuscript. AK was the second reviewer for the abstract/title screening and full text screening. BO, AK, and TYL were second reviewers for the data extraction and bias assessment verification. JEH, HYL, NMA, and LM provided critical feedback and helped shape the research, analysis, and manuscript. YAG is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgments

We acknowledge the essential help of Nia Roberts and Carolyn Smith, Bodleian Health Care Libraries, University of Oxford, UK, with my search strategy, choice of databases, and use of reference management software and systematic review software. During the course of preparing this work, the authors used ChatGPT only to improve the readability, cohesiveness, and language of the work. Following the use of this tool/service, the authors formally reviewed the content for its accuracy and edited it as necessary. The authors take full responsibility for all the content of this publication.

Conflicts of Interest

LM is a part time employee of EMIS Group plc and an advisor and shareholder of Malama Inc. The other authors declare that they have no competing interests.

References

1. H. Wang, N. Li, T. Chivese, et al., “IDF Diabetes Atlas: Estimation of Global and Regional Gestational Diabetes Mellitus Prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria,” *Diabetes Research and Clinical Practice* 183 (2022): 109050, <https://doi.org/10.1016/j.diabres.2021.109050>.
2. W. Ye, C. Luo, J. Huang, C. Li, Z. Liu, and F. Liu, “Gestational Diabetes Mellitus and Adverse Pregnancy Outcomes: Systematic Review and Meta-Analysis,” *BMJ* 377 (2022): e067946.
3. L. Bellamy, J. P. Casas, A. D. Hingorani, and D. Williams, “Type 2 Diabetes Mellitus After Gestational Diabetes: A Systematic Review and Meta-Analysis,” *Lancet (London, England)* 373, no. 9677 (2009): 1773–1779, [https://doi.org/10.1016/S0140-6736\(09\)60731-5](https://doi.org/10.1016/S0140-6736(09)60731-5).
4. B. Daly, K. A. Toulis, N. Thomas, et al., “Increased Risk of Ischemic Heart Disease, Hypertension, and Type 2 Diabetes in Women With Previous Gestational Diabetes Mellitus, a Target Group in General Practice for Preventive Interventions: A Population-Based Cohort Study,” *PLoS Medicine* 15, no. 1 (2018): e1002488, <https://doi.org/10.1371/journal.pmed.1002488>.
5. J. Lauenborg, E. Mathiesen, T. Hansen, et al., “The Prevalence of the Metabolic Syndrome in a Danish Population of Women With Previous Gestational Diabetes Mellitus Is Three-Fold Higher Than in the General Population,” *Journal of Clinical Endocrinology and Metabolism* 90, no. 7 (2005): 4004–4010, <https://doi.org/10.1210/jc.2004-1713>.
6. N. Leybovitz-Haleluya, T. Wainstock, D. Landau, and E. Sheiner, “Maternal Gestational Diabetes Mellitus and the Risk of Subsequent Pediatric Cardiovascular Diseases of the Offspring: A Population-Based Cohort Study With up to 18 Years of Follow up,” *Acta Diabetologica* 55, no. 10 (2018): 1037–1042, <https://doi.org/10.1007/s00592-018-1176-1>.
7. N. Kaseva, M. Väärasmäki, H. M. Matinolli, et al., “Pre-Pregnancy Overweight or Obesity and Gestational Diabetes as Predictors of Body Composition in Offspring Twenty Years Later: Evidence From Two

- Birth Cohort Studies,” *International Journal of Obesity* 42, no. 4 (2018): 872–879, <https://doi.org/10.1038/ijo.2017.277>.
8. H. Nijs and K. Benhalima, “Gestational Diabetes Mellitus and the Long-Term Risk for Glucose Intolerance and Overweight in the Offspring: A Narrative Review,” *Journal of Clinical Medicine* 9, no. 2 (2020): 599, <https://doi.org/10.3390/jcm9020599>.
9. E. M. Wendland, M. R. Torloni, M. Falavigna, et al., “Gestational Diabetes and Pregnancy Outcomes—A Systematic Review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) Diagnostic Criteria,” *BMC Pregnancy and Childbirth* 12 (2012): 23, <https://doi.org/10.1186/1471-2393-12-23>.
10. H. Y. Yong, Z. Mohd Shariff, B. N. Mohd Yusof, et al., “Independent and Combined Effects of Age, Body Mass Index and Gestational Weight Gain on the Risk of Gestational Diabetes Mellitus,” *Scientific Reports* 10, no. 1 (2020): 8486, <https://doi.org/10.1038/s41598-020-65251-2>.
11. D. Mennickent, A. Rodríguez, M. Fariás-Jofré, J. Araya, and E. Guzmán-Gutiérrez, “Machine Learning-Based Models for Gestational Diabetes Mellitus Prediction Before 24–28 Weeks of Pregnancy: A Review,” *Artificial Intelligence in Medicine* 132 (2022): 102378, <https://doi.org/10.1016/j.artmed.2022.102378>.
12. Z. Zhang, L. Yang, W. Han, et al., “Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis,” *Journal of Medical Internet Research* 24, no. 3 (2022): e26634, <https://doi.org/10.2196/26634>.
13. S. D. Cooray, L. A. Wijeyaratne, G. Soldatos, J. Allotey, J. A. Boyle, and H. J. Teede, “The Unrealised Potential for Predicting Pregnancy Complications in Women With Gestational Diabetes: A Systematic Review and Critical Appraisal,” *International Journal of Environmental Research and Public Health* 17, no. 9 (2020): 3048, <https://doi.org/10.3390/ijerph17093048>.
14. A. Liberati, D. G. Altman, J. Tetzlaff, et al., “The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Healthcare Interventions: Explanation and Elaboration,” *BMJ* 339, no. jul21 1 (2009): b2700, <https://doi.org/10.1136/bmj.b2700>.
15. G. J. Geersing, W. Bouwmeester, P. Zuihthoff, R. Spijker, M. Leeflang, and K. Moons, “Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews,” *PLoS ONE* 7, no. 2 (2012): e32844, <https://doi.org/10.1371/journal.pone.0032844>.
16. Z. Alfirevic, E. Keeney, T. Dowswell, et al., “Search Strategy: Cochrane Pregnancy and Childbirth Group,” *NIHR Journals Library* (2016): 46–54.
17. K. G. M. Moons, J. A. H. de Groot, W. Bouwmeester, et al., “Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist,” *PLoS Medicine* 11, no. 10 (2014): e1001744, <https://doi.org/10.1371/journal.pmed.1001744>.
18. R. Martis, C. A. Crowther, E. Shepherd, J. Alsweiler, M. R. Downie, and J. Brown, “Treatments for Women With Gestational Diabetes Mellitus: An Overview of Cochrane Systematic Reviews,” *Cochrane Database of Systematic Reviews* 2018, no. 8 (2018): (8), <https://doi.org/10.1002/14651858.CD012327.pub2>.
19. A. M. Egan, D. Bogdanet, T. P. Griffin, et al., “A Core Outcome Set for Studies of Gestational Diabetes Mellitus Prevention and Treatment,” *Diabetologia* 63, no. 6 (2020): 1120–1127, <https://doi.org/10.1007/s00125-020-05123-6>.
20. “Covidence Systematic Review Software.” www.covidence.org.
21. R. F. Wolff, K. G. M. Moons, R. D. Riley, et al., “PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies,” *Annals of Internal Medicine* 170, no. 1 (2019): 51–58, <https://doi.org/10.7326/M18-1376>.
22. B. M. Fernandez-Felix, J. López-Alcalde, M. Roqué, A. Muriel, and J. Zamora, “CHARMS and PROBAST at Your Fingertips: A Template for Data Extraction and Risk of Bias Assessment in Systematic Reviews of Predictive Models,” *BMC Medical Research Methodology* 23, no. 1 (2023): 44, <https://doi.org/10.1186/s12874-023-01849-0>.
23. “World Bank Group Country Classifications by Income Level for FY24 (July 1, 2023– June 30, 2024).” (2023). Accessed July 31, 2023. <https://blogs.worldbank.org/opendata/new-world-bank-group-country-classifications-income-level-fy24>.
24. M. W. Carpenter and D. R. Coustan, “Criteria for Screening Tests for Gestational Diabetes,” *American Journal of Obstetrics and Gynecology* 144, no. 7 (1982): 768–773, [https://doi.org/10.1016/0002-9378\(82\)90349-0](https://doi.org/10.1016/0002-9378(82)90349-0).
25. L. Jovanovic, C. M. Peterson, J. P. Daniels, et al., “International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy,” *Diabetes Care* 33, no. 3 (2010): 676–682, <https://doi.org/10.2337/dc09-1848>.
26. L. M. Harper, A. V. Glover, J. R. Biggio, and A. Tita, “Predicting Failure of Glyburide Therapy in Gestational Diabetes,” *Journal of Perinatology: The Official Journal of the California Perinatal Association* 36, no. 5 (2016): 347–351, <https://doi.org/10.1038/jp.2015.216>.
27. Y. Yuan, Q. Zhu, X. Yao, Z. Shi, and J. Wen, “Maternal Circulating Metabolic Biomarkers and Their Prediction Performance for Gestational Diabetes Mellitus Related Macrosomia,” *BMC Pregnancy and Childbirth* 23, no. 1 (2023): 113, <https://doi.org/10.1186/s12884-023-05440-9>.
28. Y. Zhang, L. Chen, L. Zhang, Y. Wu, and L. Li, “Fasting Plasma Glucose and Fetal Ultrasound Predict the Occurrence of Neonatal Macrosomia in Gestational Diabetes Mellitus,” *BMC Pregnancy and Childbirth* 23, no. 1 (2023): 269, <https://doi.org/10.1186/s12884-023-05594-6>.
29. A. Jotic, M. Stojkovic, T. Milicic, et al., “Development of ALOHa-G Risk Score for Detecting Postpartum Dyslipidemia Among Normoglycemic Women With Previous Gestational Diabetes: Observational Cohort Study,” *Diabetes Therapy* 14, no. 5 (2023): 857–867, <https://doi.org/10.1007/s13300-023-01387-4>.
30. H. McIntyre, K. Gibbons, J. Lowe, and J. Oats, “Development of a Risk Engine Relating Maternal Glycemia and Body Mass Index to Pregnancy Outcomes (Reprinted from *Diabetes Res Clin Pract*, vol 139, pg 331-8, 2018),” *Diabetes Research and Clinical Practice* 145 (2018): 31–38, <https://doi.org/10.1016/j.diabres.2018.11.010>.
31. C. Phaloprakarn and S. Tangjitgamol, “Risk Assessment for Preeclampsia in Women With Gestational Diabetes Mellitus,” *Journal of Perinatal Medicine* 37, no. 6 (2009): 617–621, <https://doi.org/10.1515/JPM.2009.108>.
32. D. A. Badr, C. Kassem, A. Carlin, et al., “Antenatal Insulin Therapy in Gestational Diabetes Mellitus: Validation of the New Brugmann Scores,” *Gynecological Endocrinology: The Official Journal of the International Society of Gynecological Endocrinology* 38, no. 5 (2022): 411–415, <https://doi.org/10.1080/09513590.2022.2048296>.
33. R. A. Barnes, T. Wong, G. P. Ross, et al., “A Novel Validated Model for the Prediction of Insulin Therapy Initiation and Adverse Perinatal Outcomes in Women With Gestational Diabetes Mellitus,” *Diabetologia* 59, no. 11 (2016): 2331–2338, <https://doi.org/10.1007/s00125-016-4047-8>.
34. R. Du and L. Li, “Estimating the Risk of Insulin Requirement in Women Complicated by Gestational Diabetes Mellitus: A Clinical Nomogram,” *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* 14 (2021): 2473–2482, <https://doi.org/10.2147/DMSO.S310866>.
35. M. Eleftheriades, C. Chatzakis, E. Papachatzopoulou, et al., “Prediction of Insulin Treatment in Women With Gestational Diabetes Mellitus,” *Nutrition & Diabetes* 11, no. 1 (2021): 1–5, <https://doi.org/10.1038/s41387-021-00173-0>.
36. H. L. Ford, I. Champion, A. Wan, M. Reddy, B. W. Mol, and D. L. Rolnik, “Predictors for Insulin Use in Gestational Diabetes Mellitus,” *European Journal of Obstetrics, Gynecology, and Reproductive Biology*

- 272, no. e41 (2022): 177–181, <https://doi.org/10.1016/j.ejogrb.2022.03.025>.
37. A. S. Koefoed, H. D. McIntyre, K. S. Gibbons, C. W. Poulsen, J. Fuglsang, and P. G. Ovesen, “Predicting the Need for Insulin Treatment: A Risk-Based Approach to the Management of Women with Gestational Diabetes Mellitus,” *Reproductive Medicine* 4, no. 3 (2023): 133–144, <https://doi.org/10.3390/reprodmed4030014>.
38. L. D. Liao, A. Ferrara, M. B. Greenberg, et al., “Development and Validation of Prediction Models for Gestational Diabetes Treatment Modality Using Supervised Machine Learning: A Population-Based Cohort Study,” *BMC Medicine* 20, no. 1 (2022): 307, <https://doi.org/10.1186/s12916-022-02499-7>.
39. H. Mendez-Figueroa, J. Daley, V. V. Lopes, and D. R. Coustan, “Predicting the Need for Medical Therapy in Patients With Mild Gestational Diabetes,” *American Journal of Perinatology* 31, no. 2 (2014): 105–112, <https://doi.org/10.1055/s-0033-1338174>.
40. D. Much, H. Jaschinski, N. Lack, et al., “Risk Stratification in Women with Gestational Diabetes According to and Beyond Current WHO Criteria,” *Hormone and Metabolic Research* 48, no. 1 (2016): 16–19, <https://doi.org/10.1055/s-0035-1565084>.
41. A. C. R. L. A. Souza, R. A. Costa, C. F. Paganoti, et al., “Can We Stratify the Risk for Insulin Need in Women Diagnosed Early With Gestational Diabetes by Fasting Blood Glucose?” *Journal of Maternal-Fetal & Neonatal Medicine: The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, and the International Society of Perinatal Obstet* 32, no. 12 (2019): 2036–2041, <https://doi.org/10.1080/14767058.2018.1424820>.
42. C. Velardo, D. Clifton, S. Hamblin, R. Khan, L. Tarassenko, and L. Mackillop, “Toward a Multivariate Prediction Model of Pharmacological Treatment for Women With Gestational Diabetes Mellitus: Algorithm Development and Validation,” *Journal of Medical Internet Research* 23, no. 3 (2021): e21435, <https://doi.org/10.2196/21435>.
43. G. Yerlikaya, V. Falcone, T. Stopp, et al., “To Predict the Requirement of Pharmacotherapy by OGTT Glucose Levels in Women with GDM Classified by the IADPSG Criteria,” *Journal Diabetes Research* 2018 (2018): 3243754, <https://doi.org/10.1155/2018/3243754>.
44. T. A. Zaccara, F. C. F. Mikami, C. F. Paganoti, R. P. V. Francisco, and R. A. Costa, “Predicting Insulin Use Among Women With Gestational Diabetes Diagnosed in Oral Glucose Tolerance Test,” *BMC Pregnancy and Childbirth* 23, no. 1 (2023): 410, <https://doi.org/10.1186/s12884-023-05746-8>.
45. B. Pintaudi, R. Fresa, M. Dalfrà, et al., “The Risk Stratification of Adverse Neonatal Outcomes in Women With Gestational Diabetes (STRONG) Study,” *Acta Diabetologica* 55, no. 12 (2018): 1261–1273, <https://doi.org/10.1007/s00592-018-1208-x>.
46. S. D. Cooray, J. A. Boyle, G. Soldatos, et al., “Development, Validation and Clinical Utility of a Risk Prediction Model for Adverse Pregnancy Outcomes in Women With Gestational Diabetes: The PerSonal GDM model,” *eClinicalMedicine* 52 (2022): 101637, <https://doi.org/10.1016/j.eclinm.2022.101637>.
47. J. S. Park, D. W. Kim, J.-Y. Kwon, Y. W. Park, Y. H. Kim, and H. Y. Cho, “Development of a Screening Tool for Predicting Adverse Outcomes of Gestational Diabetes Mellitus: A Retrospective Cohort Study,” *Medicine* 95, no. 1 (2016): e2204, <https://doi.org/10.1097/MD.0000000000002204>.
48. H. Lu, J. Hirst, J. Yang, L. Mackillop, and D. Clifton, “Standardising the Assessment of Caesarean Birth Using an Oxford Caesarean Prediction Score for Mothers With Gestational Diabetes,” *Healthcare Technology Letters* 9, no. 1-2 (2022): 1–8, <https://doi.org/10.1049/htl2.12022>.
49. S. Z. Ramos, A. K. Lewkowitz, M. G. Lord, et al., “Predicting Primary Cesarean Delivery in Pregnancies Complicated by Gestational Diabetes Mellitus,” *American Journal of Obstetrics and Gynecology* 229, no. 5 (2023): 549.e1–549.e16, <https://doi.org/10.1016/j.ajog.2023.06.002>.
50. H. S. Kim, S. Y. Oh, G. J. Cho, et al., “A Predictive Model for Large-for-Gestational-Age Infants among Korean Women With Gestational Diabetes Mellitus Using Maternal Characteristics and Fetal Biometric Parameters,” *Journal of Clinical Medicine* 11, no. 17 (2022): 4951, <https://doi.org/10.3390/jcm11174951>.
51. X. Kang, Y. Liang, S. Wang, et al., “Prediction Model Comparison for Gestational Diabetes Mellitus With Macrosomia Based on Risk Factor Investigation,” *Journal of Maternal-Fetal & Neonatal Medicine: The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, and the International Society of Perinatal Obstetrics* 34, no. 15 (2019): 2481–2490, <https://doi.org/10.1080/14767058.2019.1668922>.
52. V. A. Odinokova, V. V. Chagovets, R. G. Shmakov, et al., “Features of Serum Lipidome in Pregnant Women With Fetal Macrosomia and a Concurrence of Macrosomia With Gestational Diabetes Mellitus,” *kusherstvo i Ginekologiya (Obstetrics and Gynecology of Russia)* 2019, no. 12 (2019): 44–51, <https://doi.org/10.18565/aig.2019.12.46-53>.
53. M. Sun, B. Zhao, S. He, et al., “The Alteration of Carnitine Metabolism in Second Trimester in GDM and a Nomogram for Predicting Macrosomia,” *Journal of Diabetes Research* 2020 (2020): 4085757, <https://doi.org/10.1155/2020/4085757>.
54. P. Sun, K. Liu, X. Cui, L. Zhang, and T. Cao, “Establishment of a Nomogram Model to Predict the Risk of Macrosomia in Patients With Gestational Diabetes Mellitus,” *Journal of Maternal-Fetal & Neonatal Medicine* 36, no. 2 (2023): 2232072, <https://doi.org/10.1080/14767058.2023.2232072>.
55. T. M. Tomlinson, D. J. Mostello, K. H. Lim, J. S. Pritchard, and G. Gross, “Fetal Overgrowth in Pregnancies Complicated by Diabetes: Development of A Clinical Prediction Index,” *Archives of Gynecology and Obstetrics* 298, no. 1 (2018): 67–74, <https://doi.org/10.1007/s00404-018-4758-9>.
56. T. M. Tomlinson, A. M. Johnson, A. M. Edwards, G. A. Gross, and D. J. Mostello, “Fetal Overgrowth in Pregnancies Complicated by Diabetes: Validation of a Predictive Index in an External Cohort,” *Archives of Gynecology and Obstetrics* 303, no. 4 (2021): 877–884, <https://doi.org/10.1007/s00404-020-05768-z>.
57. Y. Yuan, Y. Li, L. Hu, and J. Wen, “Exosomal RNA Expression Profiles and Their Prediction Performance in Patients With Gestational Diabetes Mellitus and Macrosomia,” *Frontiers in Endocrinology* 13 (2022): 864971, <https://doi.org/10.3389/fendo.2022.864971>.
58. Y. Zou, Y. Zhang, Z. Yin, L. Wei, B. Lv, and Y. Wu, “Establishment of a Nomogram Model to Predict Macrosomia in Pregnant Women With Gestational Diabetes Mellitus,” *BMC Pregnancy and Childbirth* 21, no. 1 (2021): 581, <https://doi.org/10.1186/s12884-021-04049-0>.
59. J. Huang, A. M. Yeung, D. G. Armstrong, et al., “Artificial Intelligence for Predicting and Diagnosing Complications of Diabetes,” *Journal of Diabetes Science and Technology* 17, no. 1 (2023): 224–238, <https://doi.org/10.1177/19322968221124583>.
60. G. Mukherjee, J. Zhang, I. Banerjee, S. Mannan, and A. Ikomi, “Development and Validation of a New Antenatal Warning Score System (DRRiP) for Neonatal Complications of Gestational Diabetes,” *International Journal of Gynecology & Obstetrics* 162, no. 2 (2023): 730–736, <https://doi.org/10.1002/ijgo.14760>.
61. S. Hahn, S. Körber, B. Gerber, and J. Stubert, “Prediction of Recurrent Gestational Diabetes Mellitus: A Retrospective Cohort Study,” *Archives of Gynecology and Obstetrics* 307, no. 3 (2023): 689–697, <https://doi.org/10.1007/s00404-022-06855-z>.
62. N. Schwartz, M. S. Green, E. Yefet, and Z. Nachum, “Modifiable Risk Factors for Gestational Diabetes Recurrence,” *Endocrine* 54, no. 3 (2016): 714–722, <https://doi.org/10.1007/s12020-016-1087-2>.
63. A. M. Bengtson, A. L. E. Dice, M. A. Clark, R. Gutman, D. Rouse, and E. Werner, “Predicting Progression From Gestational Diabetes to

- Impaired Glucose Tolerance Using Peridelivery Data: An Observational Study," *American Journal of Perinatology* 41, no. S 01 (2022): e282–e289, <https://doi.org/10.1055/a-1877-9587>.
64. M. Kondo, Y. Nagao, M. H. Mahbub, T. Tanabe, and Y. Tanizawa, "Factors Predicting Early Postpartum Glucose Intolerance in Japanese Women With Gestational Diabetes Mellitus: Decision-Curve Analysis," *Diabetic Medicine: The Journal of the British Diabetic Association* 35, no. 8 (2018): 1111–1117, <https://doi.org/10.1111/dme.13657>.
65. V. Bartakova, B. Baratova, L. Pacal, et al., "Development of a New Risk Score for Stratification of Women with Gestational Diabetes Mellitus at High Risk of Persisting Postpartum Glucose Intolerance Using Routinely Assessed Parameters," *Life (Basel, Switzerland)* 11, no. 6 (2021): 464, <https://doi.org/10.3390/life11060464>.
66. L. Pei, F. Lai, W. Chen, et al., "Pancreatic-Derived Factor Predicts Remission of Impaired Glucose Tolerance Women With History of Gestational Diabetes," *Diabetes Research and Clinical Practice* 204 (2023): 110892, <https://doi.org/10.1016/j.diabres.2023.110892>.
67. A. A. Muche, O. O. Olayemi, and Y. K. Gete, "Predictors of Postpartum Glucose Intolerance in Women With Gestational Diabetes Mellitus: A Prospective Cohort Study in Ethiopia Based on the Updated Diagnostic Criteria," *BMJ Open* 10, no. 8 (2020): e036882, <https://doi.org/10.1136/bmjopen-2020-036882>.
68. N. Periyathambi, D. Parkhi, Y. Ghebremichael-Weldeslassie, et al., "Machine Learning Prediction of Non-Attendance to Postpartum Glucose Screening and Subsequent Risk of Type 2 Diabetes Following Gestational Diabetes," *PLoS ONE* 17, no. 3 (2022): e0264648, <https://doi.org/10.1371/journal.pone.0264648>.
69. E. R. Ingram, I. K. Robertson, K. J. Ogden, A. E. Dennis, J. E. Campbell, and A. M. Corbould, "Utility of Antenatal Clinical Factors for Prediction of Postpartum Outcomes in Women With Gestational Diabetes Mellitus (GDM)," *Australian & New Zealand Journal of Obstetrics & Gynaecology* 57, no. 3 (2017): 272–279, <https://doi.org/10.1111/ajo.12514>.
70. A. Barden, R. Singh, B. Walters, M. Phillips, and L. J. Beilin, "A Simple Scoring Method Using Cardiometabolic Risk Measurements in Pregnancy to Determine 10-Year Risk of Type 2 Diabetes in Women With Gestational Diabetes," *Nutrition & Diabetes* 3, no. 6 (2013): e72, <https://doi.org/10.1038/nu.2013.15>.
71. H. C. Lin, C. T. Su, and P. C. Wang, "An Application of Artificial Immune Recognition System for Prediction of Diabetes Following Gestational Diabetes," *Journal of Medical Systems* 35, no. 3 (2011): 283–289, <https://doi.org/10.1007/s10916-009-9364-8>.
72. B. Man, A. Schwartz, O. Pugach, Y. Xia, and B. Gerber, "A Clinical Diabetes Risk Prediction Model for Prediabetic Women With Prior Gestational Diabetes," *PLoS ONE* 16, no. 6 (2021): e0252501, <https://doi.org/10.1371/journal.pone.0252501>.
73. C. Ignell, M. Ekelund, E. Anderberg, and K. Berntorp, "Model for Individual Prediction of Diabetes up to 5 Years After Gestational Diabetes Mellitus," *Springerplus* 5 (2016): 318, <https://doi.org/10.1186/s40064-016-1953-7>.
74. A. Allalou, A. Nalla, K. J. Prentice, et al., "A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes," *Diabetes* 65, no. 9 (2016): 2529–2539, <https://doi.org/10.2337/db15-1720>.
75. M. V. Joglekar, W. K. M. Wong, F. K. Ema, et al., "Postpartum Circulating microRNA Enhances Prediction of Future Type 2 Diabetes in Women With Previous Gestational Diabetes," *Diabetologia* 64, no. 7 (2021): 1516–1526, <https://doi.org/10.1007/s00125-021-05429-z>.
76. S. R. Khan, H. Mohan, Y. Liu, et al., "The Discovery of Novel Predictive Biomarkers and Early-Stage Pathophysiology for the Transition From Gestational Diabetes to Type 2 Diabetes," *Diabetologia* 62, no. 4 (2019): 687–703, <https://doi.org/10.1007/s00125-018-4800-2>.
77. S. H. Kwak, S. H. Choi, K. Kim, et al., "Prediction of Type 2 Diabetes in Women With a History of Gestational Diabetes Using a Genetic Risk Score," *Diabetologia* 56, no. 12 (2013): 2556–2563, <https://doi.org/10.1007/s00125-013-3059-x>.
78. M. Lai, Y. Liu, G. V. Ronnett, et al., "Amino Acid and Lipid Metabolism in Post-Gestational Diabetes and Progression to Type 2 Diabetes: A Metabolic Profiling Study," *PLoS Medicine* 17, no. 5 (2020): e1003112, <https://doi.org/10.1371/journal.pmed.1003112>.
79. M. Lappas, P. A. Mundra, G. Wong, et al., "The Prediction of Type 2 Diabetes in Women With Previous Gestational Diabetes Mellitus Using Lipidomics," *Diabetologia* 58, no. 7 (2015): 1436–1442, <https://doi.org/10.1007/s00125-015-3587-7>.
80. M. Lappas, D. Jinks, A. Shub, J. C. Willcox, H. M. Georgiou, and M. Permezel, "Postpartum IGF-I and IGFBP-2 Levels Are Prospectively Associated With the Development of Type 2 Diabetes in Women With Previous Gestational Diabetes Mellitus," *Diabetes & Metabolism* 42, no. 6 (2016): 442–447, <https://doi.org/10.1016/j.diabet.2016.06.004>.
81. M. Lappas, H. M. Georgiou, A. Velagic, J. C. Willcox, M. Permezel, and A. Shub, "Do Postpartum Levels of Apolipoproteins Prospectively Predict the Development of Type 2 Diabetes in Women with Previous Gestational Diabetes Mellitus?" *Experimental and Clinical Endocrinology & Diabetes: Official Journal of the German Society of Endocrinology [Deutsche Gesellschaft für Endokrinologie] and the German Diabetes Association [Deutsche Diabetes Gesellschaft]* 127, no. 6 (2019): 353–358, <https://doi.org/10.1055/a-0577-7700>.
82. Z. Zhang, M. Lai, A. L. Piro, et al., "Intensive Lactation Among Women With Recent Gestational Diabetes Significantly Alters the Early Postpartum Circulating Lipid Profile: The SWIFT Study," *BMC Medicine* 19, no. 1 (2021): 241, <https://doi.org/10.1186/s12916-021-02095-1>.
83. W. Li, J. Leng, H. Liu, et al., "Nomograms for Incident Risk of Postpartum Type 2 Diabetes in Chinese Women With Prior Gestational Diabetes Mellitus," *Clinical Endocrinology* 90, no. 3 (2019): 417–424, <https://doi.org/10.1111/cen.13863>.
84. A. Elnour and J. McElnay, "Risk Factor Model for Developing Postpartum Diabetes Mellitus in Patients With Gestational Diabetes Mellitus," *Pharmacy World & Science* 29, no. 5 (2007): 438–439.
85. M. Kohler, A. G. Ziegler, and A. Beyerlein, "Development of a Simple Tool to Predict the Risk of Postpartum Diabetes in Women With Gestational Diabetes Mellitus," *Acta Diabetologica* 53, no. 3 (2016): 433–437, <https://doi.org/10.1007/s00592-015-0814-0>.
86. M. Marozas, S. Sosunkevic, M. Francaite-Daugeliene, D. Velickiene, and A. Lukosevicius, "Algorithm for Diabetes Risk Evaluation From Past Gestational Diabetes Data," *Technology and Health Care: The Official Journal of the European Society for Engineering and Medicine* 26, no. 4 (2018): 637–648, <https://doi.org/10.3233/THC-181325>.
87. H. Cormier, J. Vigneault, V. Garneau, et al., "An Explained Variance-Based Genetic Risk Score Associated With Gestational Diabetes Antecedent and With Progression to Pre-Diabetes and Type 2 Diabetes: A Cohort Study," *BJOG* 122, no. 3 (2015): 411–419, <https://doi.org/10.1111/1471-0528.12937>.
88. U. V. Ukah, R. W. Platt, N. Auger, K. Dasgupta, and N. Dayan, "Development and Internal Validation of a Model to Predict Type 2 Diabetic Complications After Gestational Diabetes," *Scientific Reports* 12, no. 1 (2022): 10377, <https://doi.org/10.1038/s41598-022-14215-9>.
89. C. Rodrigues, F. Cunha, F. Castro, et al., "New Clinical Score to Predict Glucose Intolerance at Postpartum Reclassification in Women With Gestational Diabetes." (2021). Accessed August 9, 2023. <https://www.easd.org/media-centre/home.html#!resources/b-new-clinical-score-to-predict-glucose-intolerance-at-postpartum-reclassification-in-women-with-gestational-diabetes-b>.
90. B. Man, A. Schwartz, Y. Xia, and B. S. Gerber, "Individualized Diabetes Risk Prediction in Women with a History of Gestational

- Diabetes,” *Diabetes* 67, no. Supplement_1 (2018): 1293, <https://doi.org/10.2337/db18-1293-P>.
91. Á. Tabák, Z. Kerényi, and G. Tamás, “Validation of an Insulin Resistance Score Calculated From Clinical Parameters in a Heterogeneous Population With Prior Gestational Diabetes,” *Diabetes Care* 23, no. 9 (2001): 1440–1445, <https://www.semanticscholar.org/paper/Validation-of-an-insulin-resistance-score-from-in-a-Tab%C3%A1k-Kere%CC%81nyi/7026352c73ab8700416e0169d1956846da561ed6>.
92. H. Kim, M. Dong, J. Lee, et al., “P08.01: A Predictive Model of Large-For-Gestational-Age Infant and Internal Validation Among Korean Women With Gestational Diabetes Mellitus: A Korean Multicentre Study,” *Ultrasound in Obstetrics & Gynecology* 52 (2018): 158–158, <https://doi.org/10.1002/uog.19676>.
93. C. D. Castilhos, R. Ladwig, M. Drehmer, B. B. Duncan, C. Facanha, and M. I. Schmidt, “Predicting Diabetes in the First Two Years Postpartum among Women with Recent Gestational Diabetes—The Linda-Brasil Study,” *Diabetes* 67, no. Supplement_1 (2018): 1447-P, <https://doi.org/10.2337/db18-1447-P>.
94. M. Caporusso, L. di Gioia, G. P. Sorice, et al., “1061-P: Need for Insulin Therapy in Gestational Diabetes Mellitus: A Predictive Model Assessment,” *Diabetes* 71 (2022): 1061-P, <https://doi.org/10.2337/db22-1061-P>.
95. O. Hourri, Y. Gil, A. Berezowsky, A. Wiznitzer, E. Hadar, and R. Chen, “339: Future Type-2 Diabetes Prediction Following Pregnancy—Using a Novel Machine Learning Algorithm,” *American Journal of Obstetrics & Gynecology* 222 (2020): S228, <https://doi.org/10.1016/j.ajog.2019.11.355>.
96. P. Šimják, K. Anderlová, A. Cinkajzlová, A. Parizek, and P. Calda, “EP20.13: A Novel Model for the Prediction of Fetal Macrosomia in Patients With Gestational Diabetes,” *Ultrasound in Obstetrics & Gynecology* 54 (2019): 363–364, <https://doi.org/10.1002/uog.21541>.
97. S. Z. Ramos, M. G. Lord, V. A. Danilack, P. Has, D. A. Savitz, and E. F. Werner, “Prediction Model for Primary Cesarean Delivery in Patients With Gestational Diabetes Mellitus,” *American Journal of Obstetrics & Gynecology* 226 (2022): S44, <https://doi.org/10.1016/j.ajog.2021.11.103>.
98. S. D. Cooray, J. A. Boyle, G. Soldatos, et al., “168-LB: Prediction for Risk-Stratified Care for Women With Gestational Diabetes: The PeRsonal GDM Study,” *Diabetes* 69, no. Supplement_1 (2020): 168-LB, <https://doi.org/10.2337/db20-168-LB>.
99. R. Barnes, T. Wong, G. P. Ross, et al., “A Model for the Prediction of Therapy Type in women with Gestational Diabetes Mellitus,” in *Proceedings of the Australasian Diabetes in Pregnancy Society Annual Scientific Meeting*, (In, ADIPS Limited, 2015).
100. M. G. Lord, P. Has, P. G. Ayers, D. A. Savitz, and M. A. Esposito, “Development and Internal Validation of a Predictive Model for Neonatal Morbidity in Pregnancies With Diabetes,” *American Journal of Obstetrics and Gynecology* 226 (2022): S218–S219, <https://doi.org/10.1016/j.ajog.2021.11.371>.
101. M. V. Joglekar, W. K. Wong, F. K. Ema, H. M. Georgiou, A. Hardikar, and M. Lappas, “94-OR: Postpartum Circulating MicroRNAs Can Predict the Development of Type 2 Diabetes in Women with Previous Gestational Diabetes,” *Diabetes* 69, no. Supplement_1 (2020): 94-OR, <https://doi.org/10.2337/db20-94-OR>.
102. C. Phaloprakarn and S. Tangjitgamol, “Risk Score for Predicting Primary Cesarean Delivery in Women With Gestational Diabetes Mellitus,” *BMC Pregnancy and Childbirth* 20, no. 1 (2020): 607, <https://doi.org/10.1186/s12884-020-03306-y>.
103. Q. F. Huang, Y. C. Hu, C. K. Wang, J. Huang, M. D. Shen, and L. H. Ren, “Clinical First-Trimester Prediction Models for Gestational Diabetes Mellitus: A Systematic Review and Meta-Analysis,” *Biological Research for Nursing* 25, no. 2 (2023): 185–197, <https://doi.org/10.1177/10998004221131993>.
104. A. J. Meehan, S. J. Lewis, S. Fazel, et al., “Clinical Prediction Models in Psychiatry: A Systematic Review of Two Decades of Progress and Challenges,” *Molecular Psychiatry* 27, no. 6 (2022): 2700–2708, <https://doi.org/10.1038/s41380-022-01528-4>.
105. L. S. Peetluk, F. M. Ridolfi, P. F. Rebeiro, D. Liu, V. C. Rolla, and T. R. Sterling, “Systematic Review of Prediction Models for Pulmonary Tuberculosis Treatment Outcomes in Adults,” *BMJ Open* 11, no. 3 (2021): e044687, <https://doi.org/10.1136/bmjopen-2020-044687>.
106. L. Daines, S. McLean, A. Buelo, S. Lewis, A. Sheikh, and H. Pinnock, “Systematic Review of Clinical Prediction Models to Support the Diagnosis of Asthma in Primary Care,” *npj Primary Care Respiratory Medicine* 29, no. 1 (2019): 19, <https://doi.org/10.1038/s41533-019-0132-z>.
107. R. D. Riley, J. Ensor, K. I. E. Snell, et al., “Calculating the Sample Size Required for Developing a Clinical Prediction Model,” *BMJ* 368 (2020): m441, <https://doi.org/10.1136/bmj.m441>.
108. R. D. Riley, G. S. Collins, J. Ensor, et al., “Minimum Sample Size Calculations for External Validation of a Clinical Prediction Model With a Time-To-Event Outcome,” *Statistics in Medicine* 41, no. 7 (2022): 1280–1295, <https://doi.org/10.1002/sim.9275>.
109. A. R. T. Donders, G. J. M. G. van der Heijden, T. Stijnen, and K. G. M. Moons, “Review: A Gentle Introduction to Imputation of Missing Values,” *Journal of Clinical Epidemiology* 59, no. 10 (2006): 1087–1091, <https://doi.org/10.1016/j.jclinepi.2006.01.014>.
110. J. Kj, D. Ar, H. Fe, et al., “Missing Covariate Data in Medical Research: To Impute Is Better Than to Ignore,” *Journal of Clinical Epidemiology* 63, no. 7 (2010): 721–727, <https://doi.org/10.1016/j.jclin.2009.12.008>.
111. Y. Vergouwe, P. Royston, K. G. M. Moons, D. G. Altman, and E. W. Steyerberg, “Development and Validation of a Prediction Model With Missing Predictor Data: A Practical Approach,” *Journal of Clinical Epidemiology* 63, no. 2 (2010): 205–214, <https://doi.org/10.1016/j.jclinepi.2009.03.017>.
112. G. Heinze, C. Wallisch, and D. Dunkler, “Variable Selection—A Review and Recommendations for the Practicing Statistician,” *Biometrical Journal* 60, no. 3 (2018): 431–449, <https://doi.org/10.1002/bimj.201700067>.
113. E. W. Steyerberg and Y. Vergouwe, “Towards Better Clinical Prediction Models: Seven Steps for Development and an ABCD for Validation,” *European Heart Journal* 35, no. 29 (2014): 1925–1931, <https://doi.org/10.1093/eurheartj/ehu207>.
114. D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, “Prognosis and Prognostic Research: Validating a Prognostic Model,” *BMJ* 338, no. may28 1 (2009): b605, <https://doi.org/10.1136/bmj.b605>.
115. B. van Calster, D. Nieboer, Y. Vergouwe, B. de Cock, M. J. Pencina, and E. W. Steyerberg, “A Calibration Hierarchy for Risk Models Was Defined: From Utopia to Empirical Data,” *Journal of Clinical Epidemiology* 74 (2016): 167–176, <https://doi.org/10.1016/j.jclinepi.2015.12.005>.
116. B. van Calster, D. J. McLernon, M. van Smeden, et al., “Calibration: The Achilles Heel of Predictive Analytics,” *BMC Medicine* 17, no. 1 (2019): 230, <https://doi.org/10.1186/s12916-019-1466-7>.
117. G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, “Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement,” *British Journal of Surgery* 102, no. 3 (2015): 148–158, <https://doi.org/10.1002/bjs.9736>.
118. K. G. M. Moons, D. G. Altman, Y. Vergouwe, and P. Royston, “Prognosis and Prognostic Research: Application and Impact of Prognostic Models in Clinical Practice,” *BMJ* 338, no. jun04 2 (2009): b606, <https://doi.org/10.1136/bmj.b606>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.