

## The unrealised potential for predicting pregnancy complications in women with gestational diabetes

Cooray, Shamil D; Wijeyaratne, Lihini A; Soldatos, Georgia; Allotey, John; Boyle, Jacqueline A; Teede, Helena J

DOI:  
[10.3390/ijerph17093048](https://doi.org/10.3390/ijerph17093048)

License:  
Creative Commons: Attribution (CC BY)

*Document Version*  
Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*  
Cooray, SD, Wijeyaratne, LA, Soldatos, G, Allotey, J, Boyle, JA & Teede, HJ 2020, '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*, vol. 17, no. 9, 3048.  
<https://doi.org/10.3390/ijerph17093048>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



Review

# The Unrealised Potential for Predicting Pregnancy Complications in Women with Gestational Diabetes: A Systematic Review and Critical Appraisal

Shamil D. Cooray <sup>1,2,\*</sup> , Lihini A. Wijeyaratne <sup>1</sup>, Georgia Soldatos <sup>1,2</sup>, John Allotey <sup>3,4</sup>,  
Jacqueline A. Boyle <sup>1,5</sup> and Helena J. Teede <sup>1,2,\*</sup>

<sup>1</sup> Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3800, Australia; lw9767.2009@my.bristol.ac.uk (L.A.W.); Georgia.Soldatos@monash.edu (G.S.); Jacqueline.Boyle@monash.edu (J.A.B.)

<sup>2</sup> Diabetes Unit, Monash Health, Clayton, VIC 3168, Australia

<sup>3</sup> Barts Research Centre for Women's Health, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AB, UK; j.allotey@qmul.ac.uk

<sup>4</sup> Multidisciplinary Evidence Synthesis Hub, Queen Mary University of London, London E1 2AB, UK

<sup>5</sup> Monash Women's Program, Monash Health, Clayton, VIC 3168, Australia

\* Correspondence: Shamil.Cooray@monash.edu (S.D.C.); Helena.Teede@monash.edu (H.J.T.)

Received: 12 March 2020; Accepted: 22 April 2020; Published: 27 April 2020



**Abstract:** Gestational diabetes (GDM) increases the risk of pregnancy complications. However, these risks are not the same for all affected women and may be mediated by inter-related factors including ethnicity, body mass index and gestational weight gain. This study was conducted to identify, compare, and critically appraise prognostic prediction models for pregnancy complications in women with gestational diabetes (GDM). A systematic review of prognostic prediction models for pregnancy complications in women with GDM was conducted. Critical appraisal was conducted using the prediction model risk of bias assessment tool (PROBAST). Five prediction modelling studies were identified, from which ten prognostic models primarily intended to predict pregnancy complications related to GDM were developed. While the composition of the pregnancy complications predicted varied, the delivery of a large-for-gestational age neonate was the subject of prediction in four studies, either alone or as a component of a composite outcome. Glycaemic measures and body mass index were selected as predictors in four studies. Model evaluation was limited to internal validation in four studies and not reported in the fifth. Performance was inadequately reported with no useful measures of calibration nor formal evaluation of clinical usefulness. Critical appraisal using PROBAST revealed that all studies were subject to a high risk of bias overall driven by methodologic limitations in statistical analysis. This review demonstrates the potential for prediction models to provide an individualised absolute risk of pregnancy complications for women affected by GDM. However, at present, a lack of external validation and high risk of bias limit clinical application. Future model development and validation should utilise the latest methodological advances in prediction modelling to achieve the evolution required to create a useful clinical tool. Such a tool may enhance clinical decision-making and support a risk-stratified approach to the management of GDM. Systematic review registration: PROSPERO CRD42019115223.

**Keywords:** gestational diabetes; prediction; risk; prognosis; pregnancy complications; large-for-gestational age; neonatal hypoglycaemia; pre-eclampsia; shoulder dystocia; systematic review

## 1. Introduction

Gestational diabetes (GDM) affects 7–20% of pregnancies and confers an increased risk of pregnancy complications with health consequences for both mother and baby. These risks are related to elevated glucose in GDM, but the relationship is complex, and an individual's risk is modified by interrelated factors, including maternal weight [1,2], gestational weight gain [3], and ethnicity [4]. Accumulating empirical data suggests this phenotypic heterogeneity may be explained by multiple physiologic defects, demonstrable on sophisticated laboratory insulin secretion and sensitivity testing [5,6]. As a result of this heterogeneity, there is a continuum and breadth in the risk of pregnancy complications associated with contemporary definitions for this condition [7,8]. Therefore, for GDM, like in much of healthcare, there is a need to move from the current one-size-fits-all approach towards a personalised and risk-stratified model-of-care.

A personalised approach would stratify women with GDM by the estimated risk of pregnancy complications. Those at high risk would maximally benefit from the targeted delivery of evidence-based preventative and therapeutic interventions. Those at low risk would be spared unnecessary treatment and may be offered less intensive intervention. Accurate risk prediction models working within existing diagnostic definitions and utilising predictors readily available in routine care, could be implementable in clinical care and would be feasible and scalable. From a public health perspective, this could enable a risk-stratified approach and development of new models of care to better allocate scarce healthcare resources, imperative in the context of the increasing GDM prevalence [9–12].

Stratifying affected women by their risk of pregnancy complications requires a method to estimate the absolute risk of future events in an individual based on readily available characteristics, or a prediction model. Many fields of medicine have seen rapid growth in the development of prediction models. However, such models are rarely translated to clinical practice [13,14], and hence rarely positively influence patient care as their creators intended. A systematic review was conducted to establish the existing literature and inform progress towards optimal primary research in prediction modelling [15] and its translation into clinical care.

The aims of this systematic review were to: identify prognostic prediction models for pregnancy complications in women with GDM; describe characteristics of the identified prognostic prediction models qualitatively; compare the performance of identified prognostic prediction models quantitatively, with meta-analysis if appropriate, and critically assess the conduct and reporting of prediction modelling development methods.

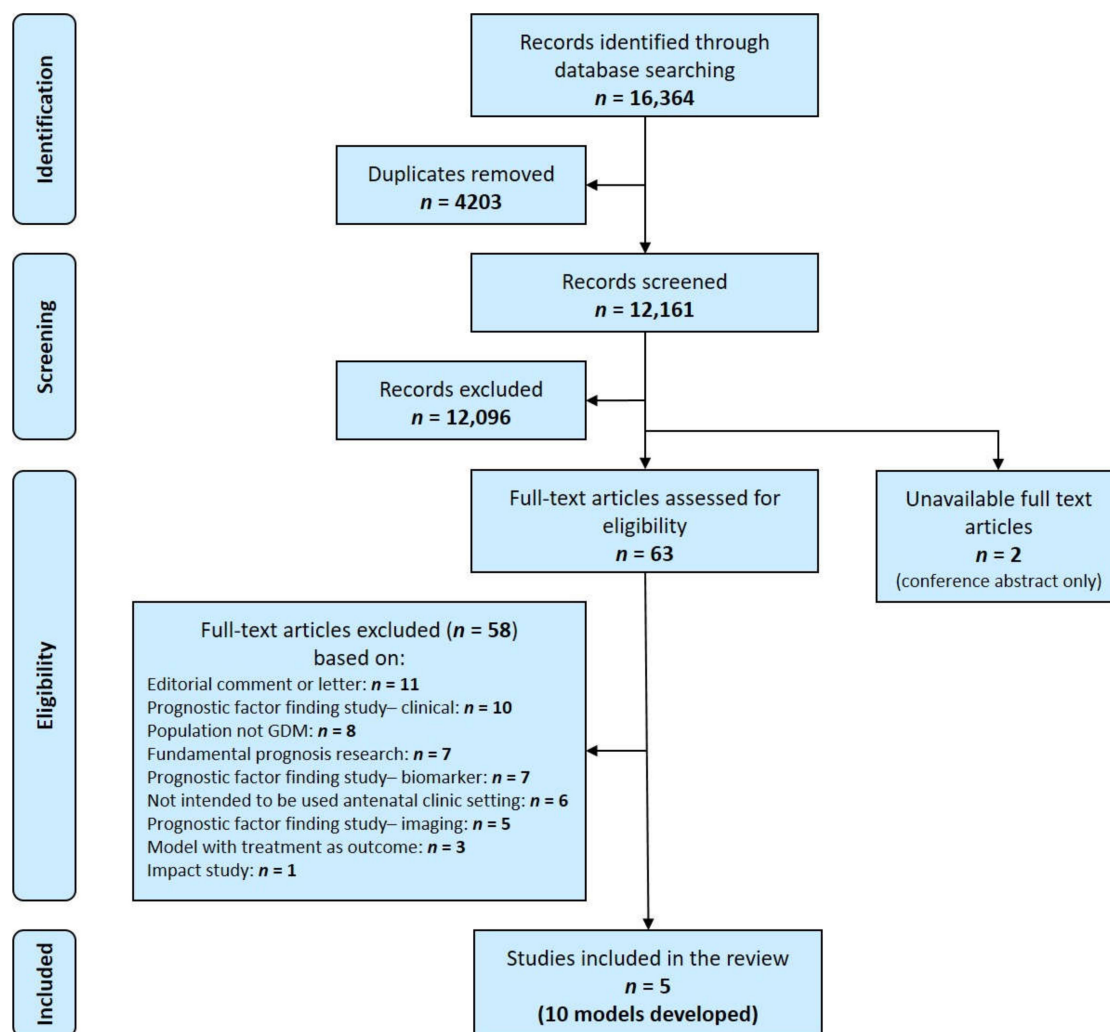
## 2. Materials and Methods

A detailed description of the methods is available in the published protocol [16]. In summary, a systematic review of prediction modelling studies for pregnancy complications in women with GDM was conducted to identify all prediction models relevant to developing a risk-stratified approach to GDM. A sensitive search strategy was developed combining the Ingui filter for prediction modelling studies [17] as updated by Geersing [18] with keywords and subject headings for gestational diabetes and relevant pregnancy complications (Supplementary Table S1). A search of MEDLINE and Embase from inception to 16 August 2019 was executed. No limits on publication date nor language were applied. Study selection, data extraction and critical appraisal were conducted independently by two reviewers. Data extraction was conducted with guidance from the CHARMS checklist (checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies) [19]. Critical appraisal focusing on the risk of bias and concerns regarding applicability was conducted using the Prediction model Risk Of Bias overdiagnosis Tool (PROBAST) [20]. The systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO), number CRD42019115223.

### 3. Results

#### 3.1. Study Selection

The search returned 12,161 unique records. Following title and abstract screening, the full text of 63 articles were assessed. Five studies meeting the selection criteria were included in this review (Figure 1) [21–25].



**Figure 1.** Flow diagram of the identification, screening and eligibility assessment of the literature for prediction models for pregnancy complications in women with gestational diabetes.

#### 3.2. Study Characteristics

The five included studies reported the development of ten prediction models for pregnancy complications in women with GDM (Table 1). No validation studies were identified. The composition of the pregnancy complications predicted varied across the studies. Four studies reported the development of a single prediction model [22–25]. One study reported the development of six models, one for each of the six outcomes (primary caesarean delivery, birth injury, large-for-gestational-age (LGA), adiposity, hyperinsulinaemia, hypoglycaemia) [21]. In this review, these six models are presented collectively due to the shared model development process and methodological characteristics.

**Table 1.** The ten models developed by the five prediction modelling studies for pregnancy complications in women with gestational diabetes.

Study	Type of Prediction Modelling Study	Model(s)
McIntyre et al. [21]	Development	A. A risk engine relating maternal glycaemia and body mass index to pregnancy outcomes with a model developed for each of: <ol style="list-style-type: none"> <li>1. Primary caesarean delivery</li> <li>2. Birth injury</li> <li>3. Large-for-gestational age</li> <li>4. Adiposity</li> <li>5. Hyperinsulinaemia</li> <li>6. Hypoglycaemia</li> </ol>
Park et al. [22]	Development	B. Screening tool for predicting adverse outcomes <sup>a</sup> of GDM
Phaloprakarn and Tangjitgamol [23]	Development	C. A risk score based on clinical characteristics of GDM women for the development of preeclampsia
Pintaudi et al. [24]	Development	D. Subgroups at different risks of developing the composite adverse neonatal outcome <sup>b</sup> by RECPAM analysis
Tomlinson et al. [25]	Development	E. The fetal overgrowth index

Abbreviations: GDM, gestational diabetes; RECPAM, RECURSIVE Partitioning and Amalgamation; <sup>a</sup> Adverse outcomes were neonatal hypoglycaemia, hyperbilirubinemia, and hyperinsulinemia; admission to the neonatal intensive care unit; large-for-gestational age; gestational insulin therapy; preeclampsia; and gestational hypertension; <sup>b</sup> Neonatal adverse outcomes were fetal growth large or small for gestational age, mortality (neonatal deaths and stillbirths), malformations, shoulder dystocia, neonatal intensive care unit need, hypoglycaemia, hypocalcaemia, hyperbilirubinemia, and respiratory distress syndrome.

### 3.3. Characteristics of the Models

#### 3.3.1. Source of Data and Participants

Four prediction modelling studies had a retrospective study design using routinely collected data, three from a single centre [22,23,25] and one from multiple centres [24] (Table 2). One prediction modelling study used a historical cohort from a single centre of a multi-centre prospective observational study [21].

The study populations varied across studies (Table 3). Two studies included only women with GDM [23,24] while three also included women without GDM [21,22,25]. Diagnostic criteria for GDM varied by region. Uniquely, one study was a post hoc analysis of a historical cohort where participants and clinicians were blinded to oral glucose tolerance test (OGTT) results [21]. Hence, in this study, 10.5% of participants who would meet the International Association of Diabetes in Pregnancy Study Groups diagnostic criteria for GDM, did not receive any specific treatment for this condition. In the other studies, treatment for GDM followed a standardised institutional protocol [23–25] or was not reported [22].

Exclusion criteria were comparable across studies with the exception of the prospective observational study with blinded OGTT results [21] (Table 3). This study had more extensive exclusion criteria to reduce the likelihood of non-adherence to the study protocol. This prediction modelling study also excluded participants with missing data for predictors and those with non-Caucasian ethnicity.

**Table 2.** Source of data and characteristics of studies used to develop models for pregnancy complications in women with gestational diabetes.

Model	Source of Data	Study Setting	Study Dates	Sample Size
A risk engine relating maternal glycaemia and body mass index to pregnancy outcomes [21]	Prospective cohort (post hoc analysis)	Single centre The antenatal clinic, Mater Misericordiae Mothers' Hospital, Australia	Jul 2000– Apr 2006	<i>n</i> = 1248
Screening tool for predicting adverse outcomes of GDM [22]	Retrospective cohort	Single centre Admitted patients, Severance Hospital, South Korea	Mar 2001– Apr 2013	<i>n</i> = 802 (306 in GDM group)
A risk score based on clinical characteristics of GDM women for the development of preeclampsia [23]	Retrospective cohort	Single centre The antenatal clinic, Vajira Hospital, Thailand	Jan 2003– Feb 2008	<i>n</i> = 813
Subgroups at different risks of developing the composite adverse neonatal outcome [24]	Retrospective cohort	Multi-centre Specialised GDM clinics, Various hospitals, Italy	Jan 2012– May 2015	<i>n</i> = 2736
The fetal overgrowth index [25]	Retrospective cohort	Single centre Collaborative diabetes in pregnancy program, South Shore Hospital, USA	Mar 2010– May 2012	<i>n</i> = 275

Abbreviation: GDM, gestational diabetes.

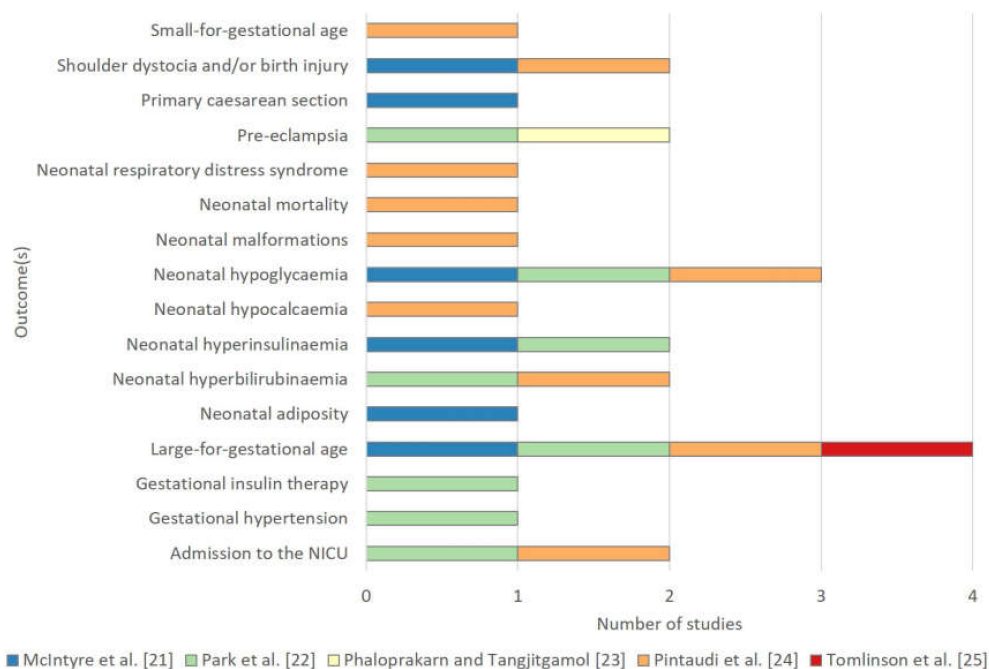
**Table 3.** Clinical characteristics of the study populations used to develop models for pregnancy complications in women with gestational diabetes.

Models	Inclusion Criteria	Exclusion Criteria	Nulliparous	Ethnicity	Diagnostic Criteria Used For GDM	Treatment Status
A risk engine relating maternal glycaemia and body mass index to pregnancy outcomes [21]	Pregnant women enrolled in HAPO study <sup>a</sup>	Multiple pregnancy, stillbirth, congenital anomaly, non-Caucasian mother, birth < 33 weeks gestation, missing data for key independent variables, age ≤ 18 years, uncertainty of gestational age, inability to complete oral glucose-tolerance test within 32 weeks gestation, assisted conception, previous diabetes requiring pharmacologic treatment or infection with hepatitis B or C virus	54.5%	100% Caucasian	NA (Participants blinded to results of OGTT)	No GDM treatment (patients and clinicians blinded to OGTT result)
Screening tool for predicting adverse outcomes of GDM [22]	Two groups: 1) Women with GDM, 2) pregnant women with false-positive glucose challenge tests	Multiple pregnancy, pre-gestational diabetes, diagnosis with GDM at <24 weeks gestation, anomalous fetuses, chronic hypertension	NR	NR	Universal screening with ACOG approach <sup>b</sup>	Treatment not reported
A risk score based on clinical characteristics of GDM women for the development of preeclampsia [23]	Women with GDM	Multiple pregnancy, risk factors for pre-eclampsia, smoking	40.0%	92% Thai, 8% South East Asian	Universal screening with ACOG approach <sup>b</sup>	GDM treatment as per standardised institutional protocol
Subgroups at different risks of developing the composite adverse neonatal outcome [24]	Women with GDM	Multiple pregnancy	45.3%	44.8% Caucasian	Risk factor-based screening with IADPSG approach <sup>c</sup>	GDM treatment as per standardised institutional protocol
The fetal overgrowth index [25]	Women with pre-gestational diabetes or GDM	Multiple pregnancy, birth <20 weeks gestation	50.5%	82% White	Universal screening with ACOG approach <sup>b</sup>	GDM treatment as per standardised institutional protocol

Abbreviations: GDM, gestational diabetes; HAPO, Hyperglycaemic and adverse pregnancy outcomes; IADPSG, International Association of Diabetes in Pregnancy Study Groups; NA, not applicable; OGTT, oral glucose tolerance test; NR, not reported; ACOG, American College of Obstetricians and Gynaecologists; <sup>a</sup> The HAPO study [26] included all pregnant women unless they had one or more exclusion criteria listed above; <sup>b</sup> ACOG approach = two-step procedure using a screening 50 g glucose challenge test (GCT) with abnormal ≥ 140 mg/dL (7.8 mmol/L) and diagnostic 100 g 3 h oral glucose tolerance test (OGTT) with two or more values above the Carpenter-Coustan cut-offs [27] considered abnormal (≥ 95 mg/dL [5.3 mmol/L] at baseline, ≥ 180 mg/dL [10.0 mmol/L] after 1 h post-load, ≥ 155 mg/dL [8.6 mmol/L] after 2 h post-load, ≥ 140 mg/dL [7.8 mmol/L] after 3 h post-load) [28]; <sup>c</sup> IADPSG approach = risk factor-based screening with a 75 g 2 h OGTT with one or more values above the IADPSG cut-offs considered abnormal (≥ 92 mg/dL [5.1 mmol/L] at baseline, ≥ 180 mg/dL [10.0 mmol/L] at 1 h post-load, ≥ 153 mg/dL [8.5 mmol/L] at 2 h post-load) [29].

### 3.3.2. Outcome(s) to be Predicted

The pregnancy complications for prediction in women with GDM varied across studies (Table 4 and Figure 2).



**Figure 2.** Outcome(s) to be predicted by studies for pregnancy complications in women with gestational diabetes.

The most common outcome to be predicted was the delivery of an LGA neonate, as it was included as an outcome in four of the five prediction modelling studies (Table 4 and Figure 2). In these four studies, LGA was defined as greater than the 90th percentile by gestational age. However, there was variation in the methods of standardization, with one study adjusting for maternal height and weight, ethnicity and parity [25] and the others adjusting for fetal sex and parity [24], fetal sex and ethnicity [22] or fetal sex alone [21]. Two of these studies included LGA neonates within a composite [22,24], it was handled as a single outcome in another study [21], and was a sole outcome for another [25].

Neonatal hypoglycaemia was selected as an outcome in three studies. It was defined as blood glucose < 2.2 mmol/L universally, however, there were variations in the timing of measures and mode of measurement were not defined in two studies [21,24]. Neonatal hyperinsulinaemia was selected as an outcome in two studies, defined as an elevated cord c-peptide level in one study [21] and as an elevated neonatal serum insulin measure in the other [24].

Two models predicted a composite outcome [22,24]. The first predicted a composite for adverse outcomes affecting both the mother and neonate [22]. The prediction of the second was limited to complications affecting the neonate [24]. However, this composite was more extensive, including 11 outcomes. Similar pregnancy complications were the subject of a third prediction modelling study, however here the outcomes were predicted in six discrete models for a single outcome rather than as a composite [21].



**Table 4.** Outcome(s) to be predicted and candidate predictors in models for pregnancy complications in women with gestational diabetes.

Model	Type	Outcome(s) to Be Predicted		Candidate Predictors		Events per Predictor
		Event	Number of Events	Number	Type	
A risk engine relating maternal glycaemia and body mass index to pregnancy outcomes [21]	Single	(1) primary caesarean delivery	241	NR	NR	–
		(2) birth injury (including shoulder dystocia)	29			–
		(3) LGA	175			–
		(4) neonatal adiposity	100			–
		(5) neonatal hyperinsulinemia	76			–
		(6) neonatal hypoglycaemia	73			–
Screening tool for predicting adverse outcomes of GDM [22]	Composite	“adverse outcomes of GDM”: neonatal hypoglycaemia, hyperbilirubinemia, and hyperinsulinemia; admission to the NICU; LGA; gestational insulin therapy; preeclampsia; gestational hypertension	458	9	demographics, patient history, physical examination, investigations	51 <sup>a</sup>
A risk score based on clinical characteristics of GDM women for the development of preeclampsia [23]	Single	preeclampsia	78	11	demographics, patient history, physical examination, investigations, disease characteristics	7
Subgroups at different risks of developing the composite adverse neonatal outcome [24]	Composite	“neonatal adverse outcome”: fetal growth large or small for gestational age, mortality (neonatal deaths and stillbirths), malformations, shoulder dystocia, NICU need, hypoglycaemia, hypocalcaemia, hyperbilirubinemia, and respiratory distress syndrome	740	7	demographics, patient history, physical examination, investigations	106
The fetal overgrowth index [25]	Single	fetal overgrowth: birthweight ≥ 90th gestational-related optimal weight centile	51	24	demographics, patient history, physical examination, investigations	2

Abbreviations: GDM, gestational diabetes; LGA, large-for-gestational age; NICU, neonatal intensive care unit; NR = not reported; <sup>a</sup> Calculated for the entire study population rather than the GDM only group.

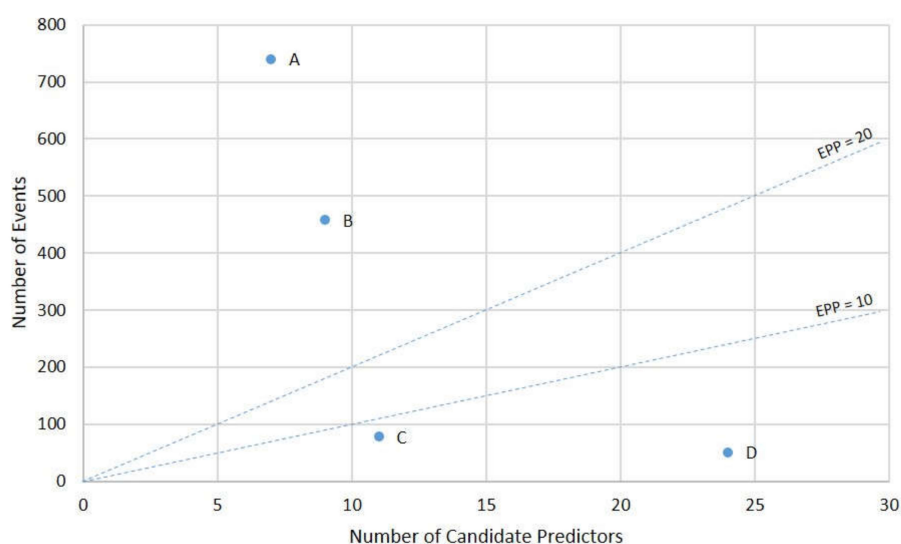
### 3.3.3. Candidate Predictors

The number of candidate predictors investigated ranged from seven to 24 (Table 4). A variety of candidate predictors were chosen for model development with a tendency towards those that are routinely available in clinical practice via patient history or physical examination, or glycaemic measures available routinely via diagnostic testing for GDM. One study considered predictors from investigations that may not be available in routine care in all settings, serum analytes (first-trimester pregnancy-associated plasma protein A and second-trimester total human chorionic gonadotropin and inhibin-A serum) and fetal abdominal circumference from an obstetric ultrasound performed between 24 and 30 weeks gestation [25].

### 3.3.4. Model Development

The presence and handling of missing data were not adequately reported in the four studies utilising routinely collected data [22–25] (Supplementary Table S2).

Statistical power varied across the studies, from two to 106 events per predictor (Table 4 and Figure 3).



**Figure 3.** Events per predictor for prediction modelling studies for pregnancy complications in women with GDM. The events per predictor (EPP) are shown, where A indicates the EPP for Pintaudi et al. [24]; B, Park et al. [22]; C, Phaloprakarn and Tangjitgamol [23]; D, Tomlinson et al. [25]. The EPP could not be calculated for McIntyre et al. [21]. An EPP above 10 to 20 is regarded as the minimum sample size for model development [19]. This graphical presentation format was adapted from Ensor and colleagues [30].

Multivariable logistic regression was the most commonly used modelling method, described in four studies [21–23,25]. Notably, a tree-based approach, namely the recursive partitioning and amalgamation (RECPAM) method, was used in one study (Supplementary Table S2) [24]. In this study a binary decision-tree that uses answers from a series of yes/no questions about clinical characteristics was developed to predict an individual’s likely outcome. Continuous predictors were dichotomised using cut-points fitted to the development data in two studies [24,25]. Methods for the selection of predictors for inclusion in and during multivariable modelling varied across the studies (Supplementary Table S2).

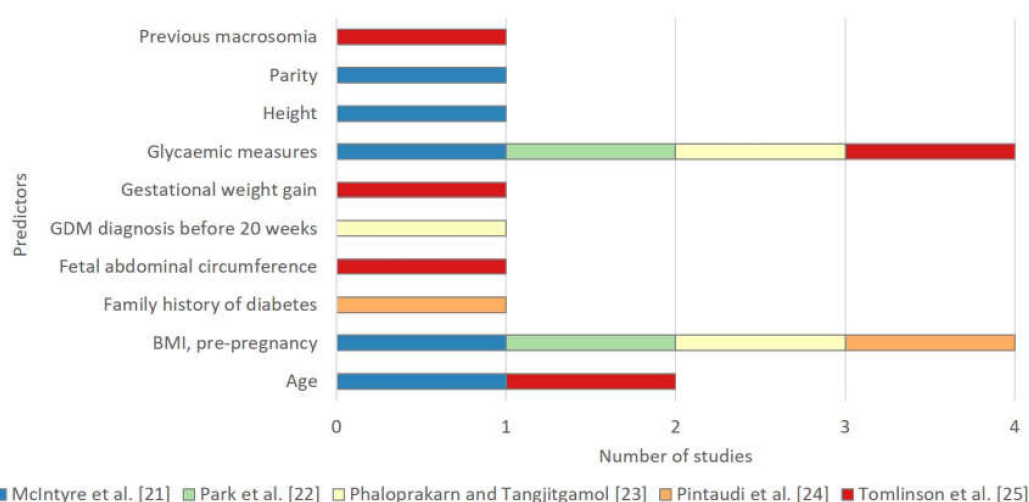
### 3.3.5. Predictors Selected in the Final Models

Ten predictors were selected for the final models across the five studies (Table 5 and Figure 4). Some measure of glycaemia and BMI was included in the final models of four of the five studies.

**Table 5.** The selected predictors, presentation format and performance of models for pregnancy complications in women with gestational diabetes.

Model	Selected Predictors	Presentation Format	Evaluation	Calibration	Performance Discrimination (c-statistic <sup>a</sup> [95% CI])
A risk engine relating maternal glycaemia and body mass index to pregnancy outcomes <sup>b</sup> [21]	Fasting, one hour, two hour OGTT results <sup>c</sup> , age, height, BMI at time of OGTT, parity	Regression coefficients without baseline components	Internal validation (apparent)	NR	Primary caesarean delivery: 0.694 (0.661–0.727) Birth injury: 0.699 (0.568–0.830) LGA: 0.654 (0.624–0.684) Adiposity 0.608 (0.570–0.645) Hyperinsulinaemia 0.601 (0.545–0.657) Hypoglycaemia 0.574 (0.517–0.632)
Screening tool for predicting adverse outcomes of GDM [22]	BMI at time of diagnosis, fasting glucose from OGTT <sup>d</sup>	Regression coefficients without baseline components	Internal validation (apparent)	GOF ( $p = 0.27$ )	0.642 (NR)
A risk score based on clinical characteristics of GDM women for the development of preeclampsia [23]	First trimester BMI $\geq 27$ kg/m <sup>2</sup> , GDM diagnosed within 20 weeks of gestation, poor glycaemic control <sup>e</sup>	Simplified scoring system	Internal validation (apparent)	GOF ( $p = 0.792$ )	0.911 (0.877–0.946)
Subgroups at different risks of developing the composite adverse neonatal outcome [24]	Pre-pregnancy BMI, family history of diabetes	Decision tree consisting of four patient subgroup classes	NR	NR	NR
The fetal overgrowth index [25]	High fasting glucose <sup>f</sup> , enlarged abdominal circumference <sup>g</sup> , excessive weight gain <sup>h</sup> , history of macrosomia <sup>i</sup> , Age $\leq 30$	Simplified scoring system	Internal validation (bootstrapping)	NR	0.89 (0.888–0.891)

Abbreviations: c-statistic, concordance statistic; CI, confidence interval, BMI, body mass index; GDM, gestational diabetes; GOF, goodness of fit; LGA, large-for-gestational-age; NR, not reported; OGTT, oral glucose tolerance test; <sup>a</sup> The concordance statistic is equal to the area under the receiver operating characteristic curve for models predicting binary outcomes; <sup>b</sup> This study presented results for each of eight fixed sets of predictors (designated models A to H) from the following: fasting, 1-hour post-load (1 h) and 2-hour post-load (2 h) OGTT results, haemoglobin A1c, age, height, BMI at time of OGTT (24–32 weeks gestation) and parity. Six of the eight sets of predictors (Models A to F) only included glycaemic measures as predictors—fasting, 1 h and 2 h glucose levels from an OGTT in varying combinations or averaged, or haemoglobin A1c. Two sets of predictors (Models G and H) combined four clinical characteristics (age, height, BMI and parity) with OGTT results, as individual components or averaged respectively. Of the eight sets of predictors evaluated the authors nominated model G (individual OGTT components with clinical characteristics) as having the best predictive ability and presented the models using this set of predictors most completely, reporting standardised coefficients for each. Hence, for comparison, we considered these models to represent the “final” model in their development process and hence, it is these results that are presented; <sup>c</sup> Results from a 75 g, 2-hour OGTT undertaken at 24 to 32 weeks gestation; <sup>d</sup> Results from a 100 g, 3-hour OGTT undertaken at 24 to 28 weeks gestation; <sup>e</sup> Poor glycaemic control defined as  $\geq 2$  separate occasions of fasting glucose  $\geq 5.8$  mmol/L and/or 2 postprandial glucose  $\geq 6.7$  mmol/L after GDM treatment; <sup>f</sup> High fasting glucose defined as fasting glucose at 24 to 30 weeks (either serum value derived from OGTT or mean fasting capillary blood glucose over 1 week)  $\geq 5.6$  mmol/L (100 mg/dL); <sup>g</sup> Enlarged abdominal circumference defined as  $\geq 90$ th percentile on ultrasound between 24 and 30 weeks; <sup>h</sup> Excessive weight gain defined as weight gain in second and third trimester  $\geq 0.3$  lb/week above the Institute of Medicine (BMI-based) goal range; <sup>i</sup> History of macrosomia defined as prior infant birthweight  $> 4$  kg.



**Figure 4.** Selected predictors in final models for pregnancy complications in women with gestational diabetes.

### 3.3.6. Model Evaluation

Model performance was evaluated using the same dataset used to develop the model (apparent validation) in all but one study [25] (Table 5). In this study, results of internal validation based on resampling of the development dataset using bootstrapping was reported. No studies reported results of external validation as a measure of transportability of the developed model to new populations.

### 3.3.7. Model Presentation

Three studies [23–25] included an alternative presentation format of the final models designed for ease of use and clinical application (Table 5).

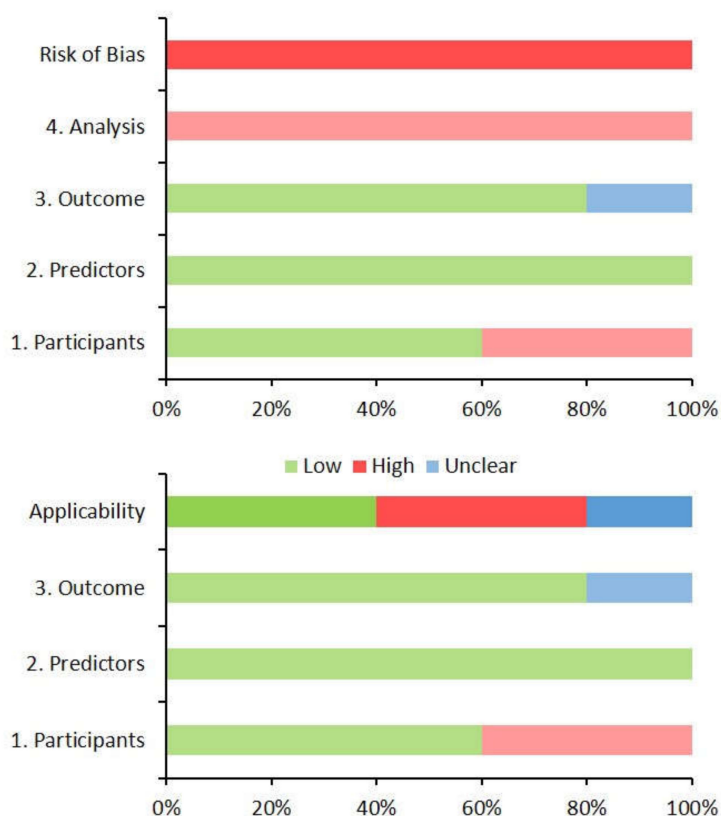
### 3.4. Comparison of Predictive Performance

Model performance was most commonly reported in terms of discrimination with four of the studies reporting a concordance statistic (*c*-statistic) for their final models [21–23,25]. For these four models predicting a binary outcome, the *c*-statistic is equal to the area under the receiver operating characteristic (ROC) curve (AUC) and ranged from 0.517 to 0.911 (Table 5). Two studies reported that calibration was adequate, presenting non-significant findings for the Hosmer–Lemeshow goodness-of-fit test [22,23]. A meta-analysis of performance measures was not appropriate because the model development studies were not sufficiently homogenous with regard to the outcome to be predicted and no validation studies for a common prediction model were identified.

Four studies reported classification measures: sensitivity and specificity and/or positive and negative predictive values [21–23,25]. Cut-points were not determined a priori. In two studies they were determined by selecting a point on the ROC curve closest to the upper left corner, which maximises the Youden index [21,22]. The method for determination was not specified in the other two studies [23,25].

### 3.5. Risk of Bias and Concerns Regarding the Applicability of Models

As assessed using the PROBAST tool [20], all models had a high risk of bias driven by the analysis domain (Figure 5 and Supplementary Table S3). There was a high concern regarding the applicability of two studies to the systematic review question [21,22], as only a minority of the participants were affected by GDM. Moreover, in one of these studies [21], the exclusion of non-Caucasian women and women with a history of gestational diabetes requiring pharmacologic treatment may further limit its applicability.



**Figure 5.** The risk of bias and concern regarding the applicability of the models developed in the five prediction modelling studies for pregnancy complications in women with gestational diabetes using the Prediction model Risk of Bias Assessment Tool (PROBAST). The *x*-axes display the proportion of studies rated by level of concern (low, high or unclear) for risk of bias or applicability for each domain.

#### 4. Discussion

This systematic review identified five prediction modelling studies for pregnancy complications in women with GDM. Approaches to prediction varied, but the birth of an LGA neonate was the leading outcome, whether as part of a composite or singularly. Models seeking to predict a single outcome were more discriminatory than those predicting a composite outcome. Three predictors emerged in most models: glycaemic measures, BMI, and maternal age. Predictive performance was generally inadequately reported, and external validation was lacking. All models had a high risk of bias due to methodologic limitations in analysis as assessed by PROBAST.

##### 4.1. Models Identified

Ten prediction models were developed by five prediction modelling studies, reflecting five distinct approaches to the clinical problem of quantifying the absolute risk of pregnancy complications in women with GDM. The literature is relatively lacking, compared to the related, but distinct literature on diagnostic prediction models for the development of GDM, with 17 models identified in a recent review [31]. However, interest in prognostic-based approaches to pregnancy risks in GDM is growing, with the first model published ten years ago and the later four within the last three years.

The model developed using a prospective cohort utilised an unselected population of pregnant women, of which 10.5% would meet current diagnostic criteria for GDM [21]. We acknowledge that the population in this prediction modelling study may not strictly meet the population criteria for eligibility of this review. However, we believe that the omission of this study would limit the value of this review, given its robust prospective study design and unique treatment-naïve study population. Furthermore, recognising that there is a continuum of risk for pregnancy complications related to GDM,

this study is valuable because it facilitates feasibility assessment for a prediction model for pregnancy complications in women with hyperglycaemia independent of the consensus-based International Association of Diabetes and Pregnancy Study Group diagnostic threshold for GDM.

A recent prediction modelling study conducted by Barnes and colleagues [32] featured prominently at the title and abstract screening stage as it seemed to be especially applicable to the review question. This study developed and externally validated a model to predict the need for insulin therapy in women with GDM. Following model development, a post-hoc analysis found that the outcome of this model, the need for insulin therapy, was strongly associated with pregnancy complications. As pregnancy complication, the outcome of interest for this review was not the subject of this prediction modelling study it was ultimately excluded. We, however, note the close relationship of these outcomes and the relevance of this existing prediction modelling study to the overarching aim of developing a stratified model-of-care for women with GDM.

#### 4.2. Characteristics of the Models

##### 4.2.1. Outcome(s) to Be Predicted

The delivery of an LGA neonate was the most common outcome predicted, offering three key advantages. Firstly, it reflects the classical maternal hyperglycaemia-fetal hyperinsulinaemia hypothesis, linking maternal hyperglycaemia to the LGA neonate, via transplacental glucose transport causing secondary fetal hyperinsulinaemia [33]. Secondly, although potentially too simplistic, excessive fetal growth is the unifying feature linking GDM to downstream pregnancy complications, such as failure to progress in labour, obstetric intervention, and shoulder dystocia. Thirdly, an LGA neonate with excess neonatal adiposity has poorer long-term metabolic health [34–38], with potential inter-generational implications [39].

Where multiple outcomes are potentially relevant, the prediction of a composite outcome, as in two of the models [22,24], may more accurately quantify multiple risks that concern women and clinicians and may be more translatable into clinical practice than a model predicting a single outcome. However, a poorly constructed composite outcome may be confusing and limit clinical application. Future model development should provide a clear rationale for the use and formulation of composite outcomes. There may be utility in heeding recommendations that the components of composite outcomes: (1) are of similar importance, (2) occur with similar frequency, and (3) are likely to have similar relative risk reductions (or predictive effects moving in the same direction) with similar underlying biology [40].

##### 4.2.2. Model Development

Two studies were inadequately powered with less than 10 events per predictor (EPP) [23,25] (Table 4 and Figure 3), generating significant risks for overfitting and consequently biased predictions [19]. The two studies which predicted composite outcome [22,24] were adequately powered for model development with more than twenty EPP [19]. This is an advantage of composite outcomes where event rates are low (Figure 4). The EPP for the six models developed in one study could not be calculated as the candidate predictors were not reported [21]. Although an EPP above 10–20 is traditionally advocated as the minimum sample size for model development [19], future studies should consider the recent proposal that a tailored sample size estimate may be advantageous in certain circumstances [41].

The dichotomisation of continuous predictor variables leads to substantial loss of information and is widely discouraged [42–45]. This is an inherent disadvantage of tree-based models included in this review [24] and was also notable in the model developed using classical regression methods [25]. In both models, continuous predictor variables were dichotomised using data-driven cut points, leading to high risks of bias [46]. In two models, continuous variables were dichotomised using cut-points which were pre-defined and independent to the development data [21,23], which leads to the loss of information but minimises the risk of bias. Continuous predictors were only handled optimally in one study [22]. Future model development studies should avoid dichotomising continuous predictors.

Selection of predictors is a key component of the model building at two stages, 1) inclusion in modelling and 2) during multivariable modelling. Where selection is based on associations with the outcome in the development dataset, there is a high risk that the developed model will be overfitted to this dataset. This was observed in three studies [22,23,25] and was inadequately reported in the other two [21,24]. In future studies, model generalisability may be improved by a priori selection of predictors for inclusion or selection independent of the predictor–outcome association, such as those based on clinical expertise [47].

#### 4.2.3. Predictors Selected in the Final Model

The predictors most commonly selected in final models, glycaemic measures ( $n = 4$ ), BMI ( $n = 4$ ), and age ( $n = 2$ ) are routinely available in clinical practice. These predictors should be included in the set of candidate predictors evaluated in future model development studies.

#### 4.2.4. Model Evaluation

Model evaluation addresses two questions pivotal for the application of prediction models into clinical practice: (1) how accurate are its predictions and (2) how generalisable is it [48]. Accuracy relates to a model's internal validity or "reproducibility" [45]. Internal validation techniques include apparent or bootstrap validation. Generalisability considers how well the model is likely to perform in a new but related population. It relates to a model's external validity or transportability, tested by evaluating the model in a new population (external validation) [45]. External validation measures predictive performance and corrects internal validation for the inherent optimism of a model being overfitted to the development dataset. As such, it provides a more realistic measure of predictive performance and quantifies the "transportability" of a model to other populations. No studies identified here reported external validation.

Three studies reported measures of model performance using apparent validation [21–23]. Here model performance is evaluated against the sample from which it was developed. The utility is limited, as performance is biased towards overestimating model performance in new populations. This risk of bias is further exaggerated with the small development datasets noted here.

One reviewed study assessed internal validity using simple bootstrapping validation, with the model repeatedly fitted to 1000 bootstrap samples and the average area under the ROC calculated as an estimate of future performance of the model in other populations [25]. This technique for internal validation is preferable to apparent validation conducted in the other studies, with less biased performance estimates. It is also preferable to split-sample validation, as data available for development and validation are maximised. Ultimately, however, external validation is essential to establish the confidence in a model required for clinical application.

The included studies did not report any formal evaluation of the clinical usefulness of the developed models. This could include the net-benefit realised by using the model's predictions to guide clinical decision-making [49]. Future studies may consider decision curve analysis [50] to quantify the clinical value of developed models as is increasingly reported in the literature [51] and recommended [52–54].

#### 4.2.5. Model Presentation

A clinical prediction model is developed with the overarching premise to be applied to clinical care to improve outcomes. Hence, developed models should be presented in a format fit for this purpose [55]. Two studies presented simplified scoring systems [23,25], and one presented a simple decision tree [24], all readily applicable to clinical care. The other studies reported regression coefficients for included predictors but did not present baseline components or regression equations [21,22], and as such do not facilitate the clinical application of the models reported.

### 4.3. Comparison of Predictive Performance

The overall performance of prediction models is traditionally quantified with two essential measures of predictive performance: discrimination and calibration [49]. Consistent with general prediction model literature [48], performance was incompletely evaluated in prediction models identified in this review (Table 5). Where reported, discrimination was evaluated using the *c*-statistic and was graphically presented with ROC curves. Calibration was evaluated using the Hosmer–Lemeshow goodness of fit test.

Discrimination was highest for models predicting single outcome versus composite outcomes overall as observed in the risk score for pre-eclampsia [23] and the fetal overgrowth index [25]. Models with single outcomes also had minimal loss in performance when adapted into simplified clinical tools. Discrimination was more limited with composite outcomes [22] and was not reported in the model developed using a decision-tree for a composite of neonatal complications [24]. In the fifth study, discrimination varied considerably for the six independent models for single pregnancy complications [21].

Calibration was evaluated in two studies using the Hosmer–Lemeshow goodness-of-fit test, suggesting an agreement between predicted and observed probabilities [22,23]. However, this method alone does not capture the magnitude or direction of miscalibration, limiting clinical utility. For instance, a model with accurate predictions in the intermediate risk range, can consistently overestimate in the low-risk range and underestimate in the high-risk range, limiting clinical usefulness. Another limitation of the Hosmer–Lemeshow test is that it is strongly related to sample size and is usually non-significant for small data sets, and conversely significant for large data sets [46]. Future work should ideally report calibration graphically (calibration plot) or in a tabular format, comparing predicted probabilities to observed outcome frequencies, and allowing the model's performance to be assessed at clinically relevant risk ranges [46,56].

### 4.4. Risk of Bias and Concerns Regarding the Applicability of Models

#### 4.4.1. Risk of Bias

Collectively, the high risk of bias for prognostic prediction models for women with GDM limits generalisability (Figure 5 and Supplementary Table S3), driven by analysis methods and overfitting of models to development datasets. This is reflective of the rapid evolution of prediction modelling methodology. The risk of bias of future prediction models may be reduced by addressing the findings of this review and referring to relevant guidelines such as the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement [57].

#### 4.4.2. Concerns Regarding Applicability

The applicability of identified models varied based on the participants in the development dataset (Figure 5 and Supplementary Table S3). Two models sought to develop prediction models for pregnancy complications as a basis for defining alternative diagnostic criteria, including women both with and without GDM [21,22]. Their aims differed from this review, the prediction of pregnancy complications in women with GDM limiting applicability.

### 4.5. Strengths and Limitations of this Review

To our knowledge, this is the first systematic review of prediction models for pregnancy complications in GDM. Strengths of this review include rigorous methods and a sensitive search strategy utilising standardised and validated search terms across the entirety of the two leading databases of biomedical literature since their inception. Bias was minimised with prospective registration and peer-reviewed publication protocol [16]. The risk of bias and applicability of included models was systematically and objectively evaluated using PROBAST [20], a robust and for-purpose tool.



Limitations included the inability to synthesize the quantitative characteristics of included models due to the heterogeneity of included studies. However, the results of the systematic search support this broad approach. We note that, despite the clinical imperative, there are only five prediction modelling studies for pregnancy complications in GDM that met our eligibility criteria in the indexed medical literature to date.

Finally, given the heterogeneity of this condition and the variety of diagnostic approaches currently used, a developed model cannot be assumed to perform equally well (or poorly) in a new, but related population. Clinical prediction modelling is an iterative multi-stage process [58]. External validation in a range of new but related populations, both geographically and temporally, is crucial but, as this review suggests, frequently neglected. Such external validation may facilitate model updating which, by addressing the characteristics of the local population and clinical practice, is likely to optimise model performance [59].

## 5. Conclusions

This review demonstrates the potential for prediction models to provide an individualised absolute risk of pregnancy complications for women affected by GDM. However, limitations in current models have been identified and this emphasises that future model development and validation would benefit from the application of methodologic advances in this rapidly evolving field. External validation, including appropriate reporting of calibration and formal evaluation of clinical usefulness with decision curve-analysis, will significantly assist the translation of promising statistical models into a useful clinical tool. Such a tool would be capable of improving outcomes for women with GDM by enhancing clinical decision-making and facilitating the stratification of affected women by their risk of pregnancy complications, thus enabling a personalised model-of-care.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1660-4601/17/9/3048/s1>, Table S1: Search strategy for MEDLINE, Table S2: Model development characteristics of models for pregnancy complications in women with gestational diabetes, Table S3: The risk of bias and concern regarding the applicability of the models developed in the five prediction modelling studies using the Prediction model Risk of Bias Assessment Tool.

**Author Contributions:** Conceptualization, S.D.C. and H.J.T.; methodology, S.D.C., H.J.T., J.A.B., G.S. and J.A.; formal analysis, S.D.C., L.A.W. and J.A.; investigation, S.D.C. and L.A.W.; data curation, S.D.C. and L.A.W.; writing—original draft preparation, S.D.C. and H.J.T.; writing—review and editing, S.D.C., H.J.T., J.A.B., G.S., L.A.W. and J.A.; visualization, S.D.C.; supervision, H.J.T., J.A.B. and G.S.; project administration, S.D.C.; funding acquisition, S.D.C. and H.J.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** S.D.C. is supported by a National Health and Medical Research Council (NHMRC) Postgraduate Scholarship, a Diabetes Australia Research Program NHMRC Top-up Scholarship, the Australian Academy of Science's Douglas and Lola Douglas Scholarship and an Australian Government Department of Education and Training Endeavour Research Leadership Award. J.A.B. is supported by a Career Development Fellowship funded by the NHMRC. H.J.T. is supported by an NHMRC Fellowship funded by the Medical Research Future Fund. The funding bodies had no role in the study design, the collection, analysis and interpretation of the data, the writing of the report, nor the decision to submit the paper for publication.

**Acknowledgments:** The authors thank Anne Young, Medical Librarian at Monash University, and Marie Misso, Head of Evidence Program at the Monash Centre for Health Research and Implementation for their expertise in developing the search strategy and systematic review methodology respectively. Preliminary findings from this work were presented at the 10th International Symposium on Diabetes, Hypertension, Metabolic Syndrome and Pregnancy (DIP 2019).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

AUC	area under the receiver operating characteristic curve
BMI	body mass index
c-statistic	concordance statistic
CHARMS	checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies
EPP	events per predictor
GDM	gestational diabetes
GWG	gestational weight gain
HAPO	hyperglycaemia and adverse pregnancy outcomes
LASSO	least absolute shrinkage and selection operator
LGA	large-for-gestational age
OGTT	oral glucose tolerance test
ROC	receiver operating characteristic
PROBAST	prediction model risk of bias overdiagnosis tool
RECPAM	recursive partitioning and amalgamation
TRIPOD	transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

## References

- Scifres, C.; Feghali, M.; Althouse, A.D.; Caritis, S.; Catov, J. Adverse Outcomes and Potential Targets for Intervention in Gestational Diabetes and Obesity. *Obstet. Gynecol.* **2015**, *126*, 316–325. [[CrossRef](#)]
- Huet, J.; Beucher, G.; Rod, A.; Morello, R.; Dreyfus, M. Joint impact of gestational diabetes and obesity on perinatal outcomes. *J. Gynecol. Obstet. Hum. Reprod.* **2018**, *47*, 469–476. [[CrossRef](#)] [[PubMed](#)]
- Goldstein, R.F.; Abell, S.K.; Ranasinha, S.; Misso, M.; Boyle, J.A.; Black, M.H.; Li, N.; Hu, G.; Corrado, F.; Rode, L.; et al. Association of Gestational Weight Gain With Maternal and Infant Outcomes: A Systematic Review and Meta-analysis. *JAMA* **2017**, *317*, 2207–2225. [[CrossRef](#)] [[PubMed](#)]
- Yuen, L.; Wong, V.W.; Simmons, D. Ethnic Disparities in Gestational Diabetes. *Curr. Diab. Rep.* **2018**, *18*, 68. [[CrossRef](#)] [[PubMed](#)]
- Powe, C.E.; Allard, C.; Battista, M.C.; Doyon, M.; Bouchard, L.; Ecker, J.L.; Perron, P.; Florez, J.C.; Thadhani, R.; Hivert, M.F. Heterogeneous Contribution of Insulin Sensitivity and Secretion Defects to Gestational Diabetes Mellitus. *Diabetes Care* **2016**, *39*, 1052–1055. [[CrossRef](#)] [[PubMed](#)]
- Benhalima, K.; Van Crombrugge, P.; Moyson, C.; Verhaeghe, J.; Vandeginste, S.; Verlaenen, H.; Vercammen, C.; Maes, T.; Dufraimont, E.; De Block, C.; et al. Characteristics and pregnancy outcomes across gestational diabetes mellitus subtypes based on insulin resistance. *Diabetologia* **2019**, *62*, 2118–2128. [[CrossRef](#)] [[PubMed](#)]
- Wan, C.S.; Abell, S.; Aroni, R.; Nankervis, A.; Boyle, J.; Teede, H. Ethnic differences in prevalence, risk factors, and perinatal outcomes of gestational diabetes mellitus: A comparison between immigrant ethnic Chinese women and Australian-born Caucasian women in Australia. *J. Diabetes* **2019**, *11*, 809–817. [[CrossRef](#)]
- Abell, S.K.; Teede, H.J. The IADPSG diagnostic criteria identify women with increased risk of adverse pregnancy outcomes in Victoria. *Aust. N. Z. J. Obstet. Gynaecol.* **2017**, *57*, 564–568. [[CrossRef](#)]
- Behboudi-Gandevani, S.; Amiri, M.; Bidhendi Yarandi, R.; Ramezani Tehrani, F. The impact of diagnostic criteria for gestational diabetes on its prevalence: A systematic review and meta-analysis. *Diabetol. Metab. Syndr.* **2019**, *11*. [[CrossRef](#)]
- Lavery, J.A.; Friedman, A.M.; Keyes, K.M.; Wright, J.D.; Ananth, C.V. Gestational diabetes in the United States: Temporal changes in prevalence rates between 1979 and 2010. *BJOG* **2017**, *124*, 804–813. [[CrossRef](#)]
- Moses, R.G.; Wong, V.C.; Lambert, K.; Morris, G.J.; San Gil, F. The prevalence of hyperglycaemia in pregnancy in Australia. *Aust. N. Z. J. Obstet. Gynaecol.* **2016**, *56*, 341–345. [[CrossRef](#)] [[PubMed](#)]
- Wong, V.W.; Lin, A.; Russell, H. Adopting the new World Health Organization diagnostic criteria for gestational diabetes: How the prevalence changes in a high-risk region in Australia. *Diabetes Res. Clin. Pract.* **2017**, *129*, 148–153. [[CrossRef](#)] [[PubMed](#)]
- Kleinrouweler, C.E.; Cheong-See, F.M.; Collins, G.S.; Kwee, A.; Thangaratinam, S.; Khan, K.S.; Mol, B.W.; Pajkrt, E.; Moons, K.G.; Schuit, E. Prognostic models in obstetrics: Available, but far from applicable. *Am. J. Obstet. Gynecol.* **2016**, *214*, 79–90 e36. [[CrossRef](#)] [[PubMed](#)]
- Wyatt, J.C.; Altman, D.G. Commentary: Prognostic models: Clinically useful or quickly forgotten? *BMJ* **1995**, *311*, 1539–1541. [[CrossRef](#)]

15. Damen, J.A.A.G.; Hooft, L. The increasing need for systematic reviews of prognosis studies: Strategies to facilitate review production and improve quality of primary research. *Diagn. Progn. Res.* **2019**, *3*, 2. [[CrossRef](#)] [[PubMed](#)]
16. Cooray, S.D.; Boyle, J.A.; Soldatos, G.; Wijeyaratne, L.A.; Teede, H.J. Prognostic prediction models for pregnancy complications in women with gestational diabetes: A protocol for systematic review, critical appraisal and meta-analysis. *Syst. Rev.* **2019**, *8*, 270. [[CrossRef](#)]
17. Ingui, B.J.; Rogers, M.A. Searching for clinical prediction rules in MEDLINE. *J. Am. Med. Inform. Assoc.* **2001**, *8*, 391–397. [[CrossRef](#)]
18. Geersing, G.J.; Bouwmeester, W.; Zuithoff, P.; Spijker, R.; Leeftang, M.; Moons, K.G. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS ONE* **2012**, *7*, e32844. [[CrossRef](#)]
19. Moons, K.G.; de Groot, J.A.; Bouwmeester, W.; Vergouwe, Y.; Mallett, S.; Altman, D.G.; Reitsma, J.B.; Collins, G.S. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med.* **2014**, *11*, e1001744. [[CrossRef](#)]
20. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; Groupdagger, P. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [[CrossRef](#)]
21. McIntyre, H.D.; Gibbons, K.S.; Lowe, J.; Oats, J.J.N. Development of a risk engine relating maternal glycemia and body mass index to pregnancy outcomes. *Diabetes Res. Clin. Pract.* **2018**, *139*, 331–338. [[CrossRef](#)] [[PubMed](#)]
22. Park, J.S.; Kim, D.W.; Kwon, J.Y.; Park, Y.W.; Kim, Y.H.; Cho, H.Y. Development of a Screening Tool for Predicting Adverse Outcomes of Gestational Diabetes Mellitus: A Retrospective Cohort Study. *Medicine (Baltimore)* **2016**, *95*, e2204. [[CrossRef](#)]
23. Phaloprakarn, C.; Tangjitgamol, S. Risk assessment for preeclampsia in women with gestational diabetes mellitus. *J. Perinat. Med.* **2009**, *37*, 617–621. [[CrossRef](#)]
24. Pintaudi, B.; Fresa, R.; Dalfra, M.; Dodesini, A.R.; Vitacolonna, E.; Tumminia, A.; Sciacca, L.; Lencioni, C.; Marcone, T.; Lucisano, G.; et al. The risk stratification of adverse neonatal outcomes in women with gestational diabetes (STRONG) study. *Acta Diabetol.* **2018**, *55*, 1261–1273. [[CrossRef](#)] [[PubMed](#)]
25. Tomlinson, T.M.; Mostello, D.J.; Lim, K.H.; Pritchard, J.S.; Gross, G. Fetal overgrowth in pregnancies complicated by diabetes: Development of a clinical prediction index. *Arch. Gynecol. Obstet.* **2018**, *298*, 67–74. [[CrossRef](#)] [[PubMed](#)]
26. Metzger, B.E.; Lowe, L.P.; Dyer, A.R.; Trimble, E.R.; Chaovarindr, U.; Coustan, D.R.; Hadden, D.R.; McCance, D.; Hod, M.; McIntyre, H.; et al. Hyperglycemia and Adverse Pregnancy Outcomes. *New Engl. J. Med.* **2008**, *358*, 1991–2002. [[CrossRef](#)]
27. Carpenter, M.W.; Coustan, D.R. Criteria for screening tests for gestational diabetes. *Am. J. Obstet. Gynecol.* **1982**, *144*, 768–773. [[CrossRef](#)]
28. Committee on Practice Bulletins-Obstetrics. ACOG Practice Bulletin No. 190: Gestational Diabetes Mellitus. *Obstet. Gynecol.* **2018**, *131*, e49–e64. [[CrossRef](#)]
29. International Association of Diabetes Pregnancy Study Groups Consensus Panel; Metzger, B.E.; Gabbe, S.G.; Persson, B.; Buchanan, T.A.; Catalano, P.A.; Damm, P.; Dyer, A.R.; Leiva, A.; Hod, M.; et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* **2010**, *33*, 676–682. [[CrossRef](#)]
30. Ensor, J.; Riley, R.D.; Moore, D.; Snell, K.I.; Bayliss, S.; Fitzmaurice, D. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. *BMJ Open* **2016**, *6*, e011190. [[CrossRef](#)]
31. Lamain-de Ruiten, M.; Kwee, A.; Naaktgeboren, C.A.; Franx, A.; Moons, K.G.M.; Koster, M.P.H. Prediction models for the risk of gestational diabetes: A systematic review. *Diagn. Progn. Res.* **2017**, *1*, 3. [[CrossRef](#)] [[PubMed](#)]
32. Barnes, R.A.; Wong, T.; Ross, G.P.; Jalaludin, B.B.; Wong, V.W.; Smart, C.E.; Collins, C.E.; MacDonald-Wicks, L.; Flack, J.R. A novel validated model for the prediction of insulin therapy initiation and adverse perinatal outcomes in women with gestational diabetes mellitus. *Diabetologia* **2016**, *59*, 2331–2338. [[CrossRef](#)] [[PubMed](#)]
33. Pedersen, J. Diabetes and Pregnancy: Blood Sugar of Newborn Infants. Ph.D. Thesis, Danish Science Press, Copenhagen, Denmark, 1952.

34. Freinkel, N. Banting Lecture 1980. Of pregnancy and progeny. *Diabetes* **1980**, *29*, 1023–1035. [[CrossRef](#)] [[PubMed](#)]
35. Lowe, W.L., Jr.; Scholtens, D.M.; Kuang, A.; Linder, B.; Lawrence, J.M.; Lebenthal, Y.; McCance, D.; Hamilton, J.; Nodzenski, M.; Talbot, O.; et al. Hyperglycemia and Adverse Pregnancy Outcome Follow-up Study (HAPO FUS): Maternal Gestational Diabetes Mellitus and Childhood Glucose Metabolism. *Diabetes Care* **2019**, *42*, 372–380. [[CrossRef](#)] [[PubMed](#)]
36. West, N.A.; Crume, T.L.; Maligie, M.A.; Dabelea, D. Cardiovascular risk factors in children exposed to maternal diabetes in utero. *Diabetologia* **2011**, *54*, 504–507. [[CrossRef](#)] [[PubMed](#)]
37. Crume, T.L.; Ogden, L.; West, N.A.; Vehik, K.S.; Scherzinger, A.; Daniels, S.; McDuffie, R.; Bischoff, K.; Hamman, R.F.; Norris, J.M.; et al. Association of exposure to diabetes in utero with adiposity and fat distribution in a multiethnic population of youth: The Exploring Perinatal Outcomes among Children (EPOCH) Study. *Diabetologia* **2011**, *54*, 87–92. [[CrossRef](#)]
38. Hammoud, N.M.; Visser, G.H.A.; van Rossem, L.; Biesma, D.H.; Wit, J.M.; de Valk, H.W. Long-term BMI and growth profiles in offspring of women with gestational diabetes. *Diabetologia* **2018**, *61*, 1037–1045. [[CrossRef](#)]
39. Gillman, M.W. Interrupting Intergenerational Cycles of Maternal Obesity. *Nestle Nutr. Inst. Workshop Ser.* **2016**, *85*, 59–69. [[CrossRef](#)]
40. Montori, V.M.; Permyer-Miralda, G.; Ferreira-Gonzalez, I.; Busse, J.W.; Pacheco-Huergo, V.; Bryant, D.; Alonso, J.; Akl, E.A.; Domingo-Salvany, A.; Mills, E.; et al. Validity of composite end points in clinical trials. *BMJ* **2005**, *330*, 594–596. [[CrossRef](#)] [[PubMed](#)]
41. Riley, R.D.; Snell, K.I.; Ensor, J.; Burke, D.L.; Harrell, F.E., Jr.; Moons, K.G.; Collins, G.S. Minimum sample size for developing a multivariable prediction model: PART II—Binary and time-to-event outcomes. *Stat. Med.* **2019**, *38*, 1276–1296. [[CrossRef](#)]
42. Dawson, N.V.; Weiss, R. Dichotomizing continuous variables in statistical analysis: A practice to avoid. *Med. Decis. Making* **2012**, *32*, 225–226. [[CrossRef](#)] [[PubMed](#)]
43. van Walraven, C.; Hart, R.G. Leave 'em alone—Why continuous variables should be analyzed as such. *Neuroepidemiology* **2008**, *30*, 138–139. [[CrossRef](#)] [[PubMed](#)]
44. Royston, P.; Altman, D.G.; Sauerbrei, W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat. Med.* **2006**, *25*, 127–141. [[CrossRef](#)] [[PubMed](#)]
45. Steyerberg, E.W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 2nd ed.; Springer International Publishing: Cham, Switzerland, 2019. [[CrossRef](#)]
46. Moons, K.G.M.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. Probast: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann. Intern. Med.* **2019**, *170*, W1–W33. [[CrossRef](#)]
47. Moons, K.G.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [[CrossRef](#)]
48. Collins, G.S.; de Groot, J.A.; Dutton, S.; Omar, O.; Shanyinde, M.; Tajar, A.; Voysey, M.; Wharton, R.; Yu, L.M.; Moons, K.G.; et al. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Med. Res. Methodol.* **2014**, *14*, 40. [[CrossRef](#)]
49. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **2010**, *21*, 128–138. [[CrossRef](#)]
50. Vickers, A.J.; Elkin, E.B. Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* **2006**, *26*, 565–574. [[CrossRef](#)]
51. Allotey, J.; Fernandez-Felix, B.M.; Zamora, J.; Moss, N.; Bagary, M.; Kelso, A.; Khan, R.; van der Post, J.A.M.; Mol, B.W.; Pirie, A.M.; et al. Predicting seizures in pregnant women with epilepsy: Development and external validation of a prognostic model. *PLoS Med.* **2019**, *16*, e1002802. [[CrossRef](#)]
52. Fitzgerald, M.; Saville, B.R.; Lewis, R.J. Decision curve analysis. *JAMA* **2015**, *313*, 409–410. [[CrossRef](#)]
53. Holmberg, L.; Vickers, A. Evaluation of prediction models for decision-making: Beyond calibration and discrimination. *PLoS Med.* **2013**, *10*, e1001491. [[CrossRef](#)] [[PubMed](#)]
54. Localio, A.R.; Goodman, S. Beyond the usual prediction accuracy metrics: Reporting results for clinical decision making. *Ann. Intern. Med.* **2012**, *157*, 294–295. [[CrossRef](#)] [[PubMed](#)]

55. Bonnett, L.J.; Snell, K.I.E.; Collins, G.S.; Riley, R.D. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* **2019**, *365*, 1737. [[CrossRef](#)]
56. Steyerberg, E.W.; Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur. Heart J.* **2014**, *35*, 1925–1931. [[CrossRef](#)] [[PubMed](#)]
57. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD statement. *Ann. Intern. Med.* **2015**, *162*, 55–63. [[CrossRef](#)] [[PubMed](#)]
58. Jansen, K.J.; Kalkman, C.J.; Grobbee, D.E.; Bonsel, G.J.; Moons, K.G.; Vergouwe, Y. The risk of severe postoperative pain: Modification and validation of a clinical prediction rule. *Anesth. Analg.* **2008**, *107*, 1330–1339. [[CrossRef](#)] [[PubMed](#)]
59. Kappen, T.H.; Vergouwe, Y.; Van Klei, W.A.; Van Wolfswinkel, L.; Kalkman, C.J.; Moons, K.G.M. Adaptation of Clinical Prediction Models for Application in Local Settings. *Med. Decis. Mak.* **2012**, *32*, E1–E10. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).