BMJ Open

Protocol for development and validation of a clinical prediction model for adverse pregnancy outcomes in women with gestational diabetes

Shamil D. Cooray [1,2] Jacqueline A. Boyle [1,3] Georgia Soldatos,[1,4] Javier Zamora [5,6] Borja M. Fernández Félix [5,7] John Allotey [8] Shakila Thangaratinam [8] Helena J. Teede [1,4]

For numbered affiliations see end of article.

Correspondence to
Prof Helena J. Teede;
Helena.Teede@monash.edu

ABSTRACT

Introduction Gestational diabetes (GDM) is a common yet highly heterogeneous condition. The ability to calculate the absolute risk of adverse pregnancy outcomes for an individual woman with GDM would allow preventative and therapeutic interventions to be delivered to women at high-risk, sparing women at low-risk from unnecessary care. The Prediction for Risk-Stratified care for women with GDM (PeRSonal GDM) study will develop, validate and evaluate the clinical utility of a prediction model for adverse pregnancy outcomes in women with GDM.

Methods and analysis We undertook formative research to conceptualise and design the prediction model. Informed by these findings, we will conduct a model development and validation study using a retrospective cohort design with participant data collected as part of routine clinical care across three hospitals. The study will include all pregnancies resulting in births from 1 July 2017 to 31 December 2018 coded for a diagnosis of GDM (estimated sample size 2430 pregnancies). We will use a temporal split-sample development and validation strategy. A multivariable logistic regression model will be fitted. The performance of this model will be assessed, and the validated model will also be evaluated using decision curve analysis. Finally, we will explore modes of model presentation suited to clinical use, including electronic risk calculators.

Ethics and dissemination This study was approved by the Human Research Ethics Committee of Monash Health (RES-19–0000713 L). We will disseminate results via presentations at scientific meetings and publication in peer-reviewed journals.

Trial registration details Systematic review proceeding this work was registered on PROSPERO (CRD42019115223) and the study was registered on the Australian and New Zealand Clinical Trials Registry (ACTRN12620000915954); Pre-results.

Strengths and limitations of this study

► We have designed a prediction model to meet an established clinical need by integrating learnings from a systematic review and critical appraisal of existing models, consensus from a clinical study steering committee and consideration of consumer perspectives.

► This study will build upon relevant literature, including a systematic review of existing prediction modelling studies to formulate a composite of prioritised, objective and serious adverse pregnancy outcomes and identify a broad series of relevant candidate predictors.

► We will adopt best practice methods for model development and validation framed by learnings from a critical appraisal of existing models.

► We will develop and validate the model using routinely-collected healthcare data in an ethnically and socioeconomically diverse population from multiple hospitals. This data was collected contemporaneously and prospectively, albeit not specifically for the purposes of this study hence missing data is likely.

► We will use decision curve analysis to formally evaluate the clinical utility of the model. This will inform the suitability of the validated model as a basis for risk-stratified model-of-care.

INTRODUCTION

Gestational diabetes (GDM) is diabetes that is first diagnosed during pregnancy, typically the second or third trimester of pregnancy and not consistent with pre-existing type 1 or type 2 diabetes.[1] It is a prominent health concern as it is common, affecting 7.5% to 27.0% of pregnancies,[2] and confers an increased risk of complications with health consequences for mother and baby.[3] However, current approaches to care are based on the false premise that the diagnostic criteria used define a group of women who are all at high-risk of adverse pregnancy outcomes.[4] In reality, the identified group is highly heterogeneous with a broad and continuous range of risk related to inter-related factors, which are inadequately integrated into the current glucocentric

treatment paradigm. Therefore, the ability to calculate the absolute risk of adverse pregnancy outcomes for an individual woman with GDM would support shared decision-making and a personalised approach to care. Here, the intensity of intervention could be stratified by risk of pregnancy complications such that preventative and therapeutic interventions could be delivered to women at high-risk, sparing women at low-risk from unnecessary intervention.

The International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria sought to translate the results of the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) study into clinical practice.[4][5] This large multinational prospective cohort study demonstrated that the risk of two adverse pregnancy outcomes (birth of a large-for-gestational-age neonate, clinical neonatal hypoglycaemia), an obstetrical intervention (primary caesarean section) and a surrogate marker for fetal hyperglycaemia (cord-blood serum C-peptide >90th percentile) was positively associated with maternal glycaemia at 24 to 28 weeks gestation as measured by an oral glucose tolerance test (OGTT). The IADPSG diagnostic criteria dichotomise the risks related to GDM on serum glucose levels using an OR of 1.75 for the above outcomes. The use of an arbitrary threshold has led to disagreement among experts and professional societies.[6][7] Indeed the optimal diagnostic strategy may vary depending on the characteristics of the local population.[1][8][9] Ultimately, these diagnostic criteria have had the unintended consequence of fostering a glucocentric approach to the treatment of GDM. This study will address this need for a more refined method of risk prediction and the targeting of intervention.

The need for refined and targeted approaches is strengthened by the heterogeneous population defined by current diagnostic criteria for GDM.[10] Pregnancy risk is clearly related to elevated glucose in GDM, but the relationship is complex, and an individual's risks are modified by interrelated factors including maternal weight,[11][12] gestational weight gain,[13] ethnicity[14] and genotype.[15] For example, it has recently been shown that within the two largest maternity services in Australia, ethnic Chinese women with GDM had a lower risk of large-for-gestational-age (LGA) babies and neonatal hypoglycaemia compared with Caucasian women, even adjusting for confounders.[16] A prediction model could integrate these risk factors to estimate risk of adverse pregnancy outcome.

The feasibility of estimating an individual's absolute risk of adverse pregnancy outcomes by integrating oral glucose tolerance test results, maternal weight and pregnancy history was established in our systematic review.[17] However, critical appraisal established that existing prediction models were not yet suitable for application to clinical practice due to high risks of bias due to methodological limitations.

The Prediction for Risk-Stratified care for women with GDM (PeRSonal GDM) study will leverage the rapidly evolving methodological advances in prediction modelling to achieve the evolution required to transform promising statistical models into useful clinical tools. In this project, we integrate the findings of this systematic review and critical appraisal of existing models, pertinent findings from landmarks trials, clinical expertise and best practice methods from contemporary guidelines to inform the methodological design of the PeRSonal GDM study.

Objectives

The aims of the PeRSonal GDM study are to:
1. Develop and internally validate a prediction model for adverse pregnancy outcomes in GDM to aid shared decision-making and stratify care;
2. Externally validate the model to demonstrate temporal transportability;
3. Evaluate the clinical utility of the model as a basis for a risk-stratified model-of-care.

METHODS AND ANALYSIS
Prediction model design

We conducted formative research to conceptualise and design a robust and clinically acceptable prediction model. First, a systematic review and critical appraisal of existing prediction models for adverse pregnancy outcomes in women with GDM was conducted following a peer-reviewed protocol.[18] Second, the study steering committee comprising two obstetricians, three endocrinologists and a neonatologist formulated key clinical requirements of the prediction model (table 1). A model addressing these requirements was designed (figure 1). Finally, a multidisciplinary clinical working group was formed to provide feedback on the proposed requirements, gauge its clinical acceptability and consider its clinical application. The working group included endocrinologists (n=9), diabetes nurse educators (n=3), dieticians (n=2), midwives (n=2), administration staff (n=2) and an obstetrician (n=1) actively involved in the provision of GDM care at several maternity hospitals. We considered consumer perspectives throughout this process, from parallel qualitative research on GDM diagnosis and risk.[19]

Study design

We will conduct a prediction model development and validation study using a retrospective cohort design. It will be conducted following expert guidance for model development and validation,[20–25] and reported per the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement.[26]

Data sources and validation strategy

This study will use routinely collected health data for pregnancies resulting in a birth from 1 July 2017 to 31 December 2018 from an existing pregnancy outcomes database from a maternity service. Maternal, obstetrical

Table 1 The fundamental requirements of a prediction model for adverse pregnancy outcomes in women with gestational diabetes

Criteria	Specifications
(1) Prognostic versus diagnostic prediction model	The aim is to predict future events (prognostic prediction model)
(2) Intended scope	To inform clinicians' therapeutic decision-making and serve as a rational basis for the stratification of GDM care
(3) The target population to whom the prediction model applies	Pregnant women with GDM, per diagnostic criteria in clinical practice
(4) The outcome to be predicted	Pregnancy complications related to GDM affecting the mother (obstetrical or maternal) or the baby (fetal or neonatal)
(5) Timespan of prediction	Complications occurring during pregnancy or soon after birth
(6) Intended moment of using the model	At diagnosis of GDM, typically at 24 to 28 weeks gestation but may be earlier

Framework adapted from that originally proposed by Moons and colleagues to consider in framing a systematic review of prediction modelling studies.[48]
GDM, gestational diabetes.

and neonatal data are collected prospectively for all women booked to deliver their baby at the service. This data is collected with consent as part of routine clinical care. This data is of high-quality and completeness as it is collected under statute with the primary aim to facilitate improvements in quality of care. We will link these data deterministically to pathology data and clinical data extracted from the medical record of the parent health service. Linked pathology data is available for approximately 70% of pregnancies, and linked clinical data is available for approximately 90% of pregnancies. All collected data will be rendered non-identifiable for all research purposes, including analysis.

The data will be split by time into two groups (analysis type 2b in TRIPOD).[27] We will develop the prediction model using pregnancies resulting in births from the first 12 months of the study period (1 July 2017 to 30 June 2018). Pregnancies resulting in births from the last 6 months of the study period (1 July 2018 to 31 December 2018) will be used to evaluate the predictive performance of the developed model (external validation). This strategy will evaluate the temporal transportability of the model.

Participants

Study setting
This maternity service is one of the largest in Australia, provides universal access to healthcare comprising multiple large maternity hospitals and serves an ethnically and socioeconomically diverse population within a catchment of 1.6 million in South-East Melbourne. All
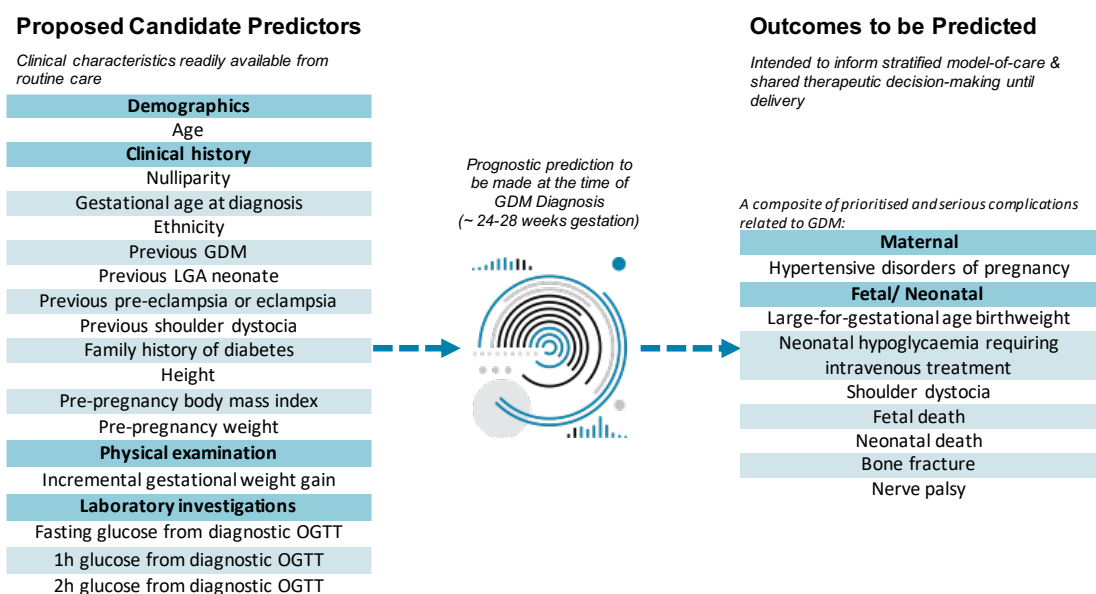

Figure 1 The design of the PeRSonal Pregnancy GDM Risk Model—Prediction for Risk-Stratified care for women with GDM. GDM, gestational diabetes; IV, intravenous; LGA, large-for-gestational-age; OGTT, oral glucose tolerance test.

levels of maternity care are available across the three hospitals with shared staff and institutional protocols and practices. Maternity care is provided to more than 9000 women each year.

Eligibility criteria

Pregnancies coded for GDM during the study period stated above will be included. There will be no exclusion criteria.

Treatment received

GDM is diagnosed and treated following institutional protocol and practices. At our service GDM is diagnosed using the International Association of Diabetes and Pregnancy Study Groups 2010 criteria,[4] as endorsed by the Australian Diabetes in Pregnancy Society with universal screening at 24 to 28 weeks with a one-step procedure using the 75 g OGTT.[6] Early screening is based on the presence of risk factors as soon as practicable using the same testing procedure with a repeat at 24 to 28 weeks if negative. The treatment package for GDM consists of an initial 2-hour group education session with diabetes nurse educator and dietician. Lifestyle management involves dietary modification, physical activity and weight management. Follow-up reviews occur with an endocrinologist or endocrinology specialist trainee every 1 to 3 weeks. Insulin is commenced where glucose targets (fasting <5.5 mmol/L and 2-hour post-prandial <7.0 mmol/L) are not met and are not amenable to further dietary modification. Metformin is used where there is evidence of significant insulin resistance, where targets are not achieved with insulin alone or when insulin use is relatively contraindicated due to the risk of significant psychological harm.

Outcome

The outcome to be predicted will be a composite consisting of a combination of eight prioritised, objective and serious adverse pregnancy outcomes defined in table 2.

Formulation of outcome(s) to be predicted

The study steering committee considered a large number of adverse pregnancy outcomes for inclusion in the composite (online supplemental table S1). Outcomes predicted by existing models identified in our systematic review and predicted by a related model for insulin therapy initiation[28] were considered. The committee also considered outcomes in the final core outcome set (COS) for GDM treatment research.[29] Reference to the COS for future GDM treatment research provided objective prioritisation of outcomes from a large international multidisciplinary group of relevant stakeholders. Finally, the committee considered all outcomes studied in the HAPO study,[5] the landmark international multicentre observational study that demonstrated associations between increasing levels of glucose levels on oral glucose tolerance testing and adverse pregnancy outcomes. From this, a composite outcome was constructed to reflect the

Table 2 The adverse pregnancy outcomes to be predicted: definition, variable type and categories

Outcome	Definition
Maternal	
Hypertensive disorders of pregnancy	Pregnancy-induced hypertension, pre-eclampsia or eclampsia
Fetal/neonatal	
LGA	Birth weight >90th percentile corrected for gestation and fetal sex using Australian population growth chart[56]
Neonatal hypoglycaemia requiring intravenous treatment	A neonate with a low blood glucose level fulfilling institutional criteria for intravenous treatment consisting of either a dextrose bolus or dextrose infusion
Shoulder dystocia	When, after delivery of the head, the baby's anterior shoulder gets caught above the mother's pubic bone
Fetal death	Death of fetus after 20 weeks gestation
Neonatal death	Death of live-born neonate
Bone fracture	Neonatal fracture (femur, humerus, clavicle or skull) suffered at birth
Nerve palsy	Neonatal nerve palsy (brachial plexus injury or facial nerve injury) suffered at birth

LGA, large-for-gestational-age.

multiple adverse pregnancy outcomes related to GDM. Construction of the composite outcome considered recommendations that components are (1) 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.[30] The rationale for inclusion or exclusion from the composite outcome to be predicted is presented in online supplemental table S2.

Outcome assessment

LGA assessment will be based on a population-based growth chart rather than customised centiles to avoid incorporation of predictor information such as ethnicity into outcome assessment. Blinding to predictors in the assessment of the outcome will not be feasible.

Predictors

Definition of predictors and measurement

Candidate predictors to be evaluated for inclusion in the model are defined in table 3. There will be no blinding between the assessment of a predictor and the outcome nor to other predictors.

Identification of candidate predictors

Candidate predictors were identified from those selected for the final models included in the systematic review

Table 3 Candidate predictors to be evaluated in model development: definition, variable type and units/ categories

Candidate predictor	Definition	Variable type	Units/categories
Demographics			
Age	Mother's age	Continuous	years
Clinical history			
Nulliparity	The condition in a woman of never having given birth	Binary	0 'No' 1 'Yes'
Gestational age at diagnosis	Gestational age at diagnosis of GDM in the index pregnancy	Continuous	weeks' gestation
Ethnicity	Self-reported ethnicity with classification aligned to the Australian Standard Classification of Cultural and Ethnic Groups[57]	Categorical	Ethnicity classified into approximately five to six categories
Previous GDM	Previous diagnosis of GDM	Binary	0 'No'; 1 'Yes'
Previous LGA	Previous child with birthweight >90th percentile corrected for gestation and fetal sex using Australian population growth chart[56]	Binary	0 'No' 1 'Yes'
Previous pre-eclampsia or eclampsia	Pre-eclampsia or eclampsia in a previous pregnancy	Binary	0 'No' 1 'Yes'
Previous shoulder dystocia	Shoulder dystocia in a previous pregnancy	Binary	0 'No' 1 'Yes'
Family history of diabetes	Any family history of diabetes	Binary	0 'No' 1 'Yes'
Height	The mother's self-reported height at about the time of conception.	Continuous	centimetres (cm)
Body mass index	Body mass divided by the square of the body height	Continuous	kg/m2
Weight	Mother's self-reported weight (body mass) about the time of conception	Continuous	kilograms (kg)
Physical examination			
Incremental gestational weight gain	Weight at first GDM clinic appointment (at around 30 weeks gestation) minus preconception weight divided by gestational weeks completed at the time of the first GDM clinical appointment	Continuous	kg
Laboratory investigations			
Fasting glucose from diagnostic OGTT	Glucose level from baseline or time zero of diagnostic oral glucose tolerance test	Continuous	mmol/L
1-hour glucose from diagnostic OGTT	Glucose level 1-hour following a 75 g oral glucose load of diagnostic oral glucose tolerance test	Continuous	mmol/L
2-hour glucose from diagnostic OGTT	Glucose level 2-hour following a 75 g oral glucose load of diagnostic oral glucose tolerance test	Continuous	mmol/L

BMI, body mass index; GDM, gestational diabetes; LGA, large-for-gestational-age; OGTT, oral glucose tolerance test.

of models for pregnancy complications in women with GDM, selected in a model for GDM diagnosis previously developed by our group,[31] and selected in a related model for insulin therapy initiation.[28] (online supplemental table S3) From these existing related models 13 of the 16 predictors will be evaluated for inclusion in this prediction modelling study (table 3). Three predictors selected for related models (poor glycaemic control, enlarged abdominal circumference and HbA1c (glycatedhaemoglobin) at diagnosis) could not be evaluated in this study as the data are not routinely collected at our service.

One previous study selected history of macrosomia as a predictor for LGA.[32] Indeed, in clinical practice, past history is often seen as a major risk factor for future occurrence. Therefore, this study will evaluate previous histories of components of the composite outcome for

inclusion in the model. Such data is available for macrosomia, LGA, pre-eclampsia and eclampsia, and shoulder dystocia, and therefore, these four predictors will be evaluated as candidate predictors.

In addition to the candidate predictors identified from their use in existing related models, ethnicity and gestational weight gain (GWG) were identified as potential predictors requiring formal evaluation due to the emergence of evidence supporting their role as significant prognostic factors. Chinese women affected by GDM were at a lower risk of a range of adverse pregnancy outcomes including LGA and neonatal hypoglycaemia compared with affected Caucasian women in an Australian cohort,[16] and South Asian babies exposed to GDM were smaller across gestation than babies of White European in an English cohort.[33] Emerging physiological data

suggests highly variable degrees of beta-cell function and insulin resistance among women diagnosed with GDM,[34] and that classifying women with GDM by these physiological defects may stratify women by their risk of adverse pregnancy outcomes.[35] Ethnicity may serve as a surrogate marker for these physiological defects avoiding the need for additional investigations. Hence, ethnicity is an appealing candidate predictor for models to predict the development of adverse pregnancy outcomes.

GWG has also been shown to be a risk factor for adverse pregnancy outcomes, independent of body mass index (BMI).[13] Specifically, GWG is associated with an increased proportion of LGA over and above that which is associated with GDM and overweight or obesity, in a general obstetric population.[36] BMI, parity and GWG together, better predict adverse pregnancy outcomes than BMI alone in a cohort attending a general antenatal clinic (women with GDM and normoglycaemia).[37] The effect of GWG is likely to be modified by other predictors, including ethnicity, supporting its integration within a multivariable model rather than a single prognostic factor-based approach.

Data extraction

We will extract records for eligible participants to create a research data set with each observation representing a pregnancy. Participants may be included more than once due to multiple pregnancy or repeat pregnancies within the study period. We will manually review eligible participant's medical record to ensure the accuracy of the diagnosis of GDM. Linked pathology and additional clinical data will be extracted and merged with the research data set. The research data set will be rendered non-identifiable for all subsequent analyses.

Sample size

In this study, the adequacy of the sample size of our developmental data set will be determined by the total number of events of the composite binary outcome. Approximately 9000 women are delivered annually at the institution from which the development data set will be derived. The prevalence of GDM at this institution is 18% (unpublished data). Therefore, over the 12-month period used for model development, we conservatively estimate that the development data set will include 1620 cases of women with GDM. We anticipate that at least 10% of these women will deliver neonates that have a birth weight, that is, LGA defined as greater than the 90th percentile for the population (approximately 162 events). Furthermore, using unpublished data from our institution, the prevalence of hypertensive disorders of pregnancy is 7% (approximately 113 events) and neonatal hypoglycaemia requiring intravenous treatment is 11% (approximately 178 events). Therefore the expected event count is greater than 453 once the additional contribution of the less common component outcomes are also considered (shoulder dystocia, fetal death, neonatal death, bone fracture, nerve palsy). Given we envisage including up to

20 candidate predictors, our study should be adequately powered as the data set will have in excess of 10 events per predictor as is commonly recommended to avoiding overfitting.[38]

Over the 6-month period used for external validation, the expected event count is 50% of that for the 12-month period used for development, hence approximately 225. This is greater than the recommended minimum of 100 events for validation.[39]

Missing data

We do not expect considerable missing data, but some will inevitably occur, with not all cases providing all variables of interest. Handling of missing data will be determined individually on a per predictor basis. The missing indicator method will be used for predictors where data is missing not at random. Multiple imputation by chained equations will be used to impute missing data as long as the data is missing at random. If necessary, we will include a supplementary table comparing predictor distributions between patients with missing data and patients with complete data.

Statistical analysis methods

To make individualised predictions for the binary composite of an adverse pregnancy outcome, we will apply a logistic regression model with the composite outcome as the dependent variable.

Handling of predictors

Continuous variables will be kept as continuous in the model (rather than dichotomising), to avoid a loss of prognostic information. Those predictors that are highly correlated with others contribute little information and will be excluded from the statistical analysis.

The functional form of the relationship of continuous predictors with the outcome will be assessed. If non-linear they will be modelled with fractional polynomials (FP). If this is the case, as several continuous variables were included in the model, we will use the multivariable fractional polynomial algorithm. Multiple imputation and FPs will be combined using the procedure described by Morris and colleagues.[40]

Model-bBuilding procedures (including predictor selection)

Candidate predictor variables will be selected a priori based on existing literature and clinical expertise as described above. During modelling, predictors will be selected by using a LASSO (Least Absolute Shrinkage and Selection Operator) method, which simultaneously selects the variables and penalises the model coefficients for over-optimism.[41]

Examination of predictor interactions will be undertaken for the following groups of predictors: weight, GWG and BMI, and fasting, 1-hour and 2-hour glucose levels from OGTT.

Internal validation and assessment of model performance

The model performance will be assessed in terms of discrimination and calibration. We will use a bootstrap

re-sampling technique to adjust for over-optimism in the estimation of model performance due to validation in the same data set that is used to develop the model itself. We will use the area under the curve of the receiver operating characteristic curve with 95% CI to assess the overall discriminatory ability of the developed model. We will report the apparent and adjusted for over-optimism model performance. A calibration plot will be created. This plot will facilitate the graphical assessment of calibration by putting affected women into groups ordered by predicted risk and considering the agreement between the mean predicted risk and the observed events in each risk group, usually deciles. The calibration will be summarised using the intercept and slope of the calibration plot. Internal validation, where the model's predictions are compared with the observed data, should return perfect calibration to the development data (calibration slope=1).

External validation

External validation of the developed model will be undertaken to assess temporal transportability. It will be undertaken using the model coefficients from the developed model to calculate the risk for each woman. We will report the predictive performance in a more recently treated cohort at the same maternity service using the same measures of discrimination and calibration as used in internal validation. Development and validation data are identical in terms of eligibility criteria, outcome and predictors.

Presentation of a simplified model for clinical use

Once a final model is identified, we will simplify and adapt the presentation of the model to facilitate its application to clinical practice. Alternative modes of presentation will be explored with a focus on maximising end-user usability and promoting translation into clinical care. Various presentation formats will be considered, including a simplified scoring system, nomogram and web-based or application-based electronic risk calculators.

Assessment of clinical utility

To supplement traditional measures of predictive model performance, discrimination and calibration, clinical utility will be formally evaluated. We will use decision curve analysis to explore the net benefit of developed models over the entire range of probability thresholds.[23 27 42] We will represent the net benefit as a function of the decision threshold in a decision curve plot. This will explore whether there is an overall net-benefit for using the models to stratify the population into two risk groups as a basis for a risk-stratified model of care:

1. Low-risk where the risk of adverse pregnancy outcomes is less than a pre-specified value—this group may be considered for a less intensive model-of-care;
2. High-risk where the risk is greater than a pre-specified value—this group should receive specialist-led hospital-based care.

Further formative research is planned to ascertain optimal risk thresholds. This will include engagement with stakeholders, including women affected by GDM and clinicians. A combination of focus groups and an electronic survey will be used.

Sensitivity analyses

We will conduct additional analysis to address the confounding effect of insulin treatment on predictor-outcome associations and hence the performance of the prediction model. This will consider four possible approaches with sensitivity analysis used to evaluate the robustness of each:

1. Derivation of a propensity score of being treated with insulin based on women pre-treatment characteristics. We will then weight observations by using the inverse probability of treatment weighting (IPTW). In this way, women with lower propensity to be treated will have more weight in the development of the prognostic model than those who had a higher probability of being treated.
2. Inclusion of insulin treatment as a component of the composite outcome.
3. Exclusion of cases where insulin treatment was used.
4. Exploration of the multinomial regression model framework for combinations of the composite outcome of adverse pregnancy outcome and insulin treatment.

The primary analysis will develop and validate a model based on clinical characteristics. Prognosis may also be influenced by an affected woman's capacity to implement lifestyle measures such a dietary modification and increased exercise. Therefore, we will undertake a sensitivity analysis to evaluate whether measures of socioeconomic disadvantage can improve the prediction of adverse pregnancy outcomes.

All statistical analysis will be performed using Stata V.16.1 (College Station, Texas: StataCorp LLC).

Patient and public involvement

No patient and public involvement in the development of this protocol. Patient and public perspectives will be essential to the formative research required to implement findings of this model development and validation study into clinical practice. As such patients and public will be invited to participate in this phase of our research.

DISCUSSION

Strengths

The formative research undertaken established the clinical need for a robust prediction model for adverse pregnancy outcomes in GDM to support therapeutic decision-making and stratification of care. Engagement with stakeholders in the model design stage should improve the clinical acceptability of the model and support future implementation efforts. The composite outcome of prioritised, objective and serious adverse events was formulated with reference to a systematic

review and critical appraisal of existing models (manuscript submitted for publication, 2020), the relevant core outcome set[43] and clinical expertise of endocrinologists, obstetricians and a neonatologist. This composite will be composed of LGA, neonatal hypoglycaemia, hypertensive disorders of pregnancy, shoulder dystocia, severe birth trauma (nerve palsy and bone fracture) and perinatal death. The transportability of the developed model will also be enhanced by the selection of candidate predictors using existing literature and clinical expertise, independent of the predictor-outcome association in the development data set.

Prediction of a composite outcome will more accurately quantify the multiple adverse pregnancy outcomes related to GDM and therefore, will be more translatable into clinical practice. This composite will be valid and clinically useful because the component outcomes are of similar importance, the three main components (LGA, neonatal hypoglycaemia and hypertensive disorders of pregnancy) occur with a similar frequency (approximately 10%),[44] and the predictive effects are likely to move in the same direction due to similar underlying biology.[30]

A method to estimate the absolute risk of adverse pregnancy outcomes for an individual woman affected by GDM would be of great benefit to affected woman, their clinicians and the health system. It would allow affected woman to better understand the implication of GDM on their pregnancy and facilitate shared-decision making with clinicians regarding the relative risks and benefits of interventions. At a system-level these individualised risk estimates would support a risk-stratified model-of-care which recognises the breadth and continuum of pregnancy risk attributable to GDM such that preventative and therapeutic interventions could be delivered to women at high-risk, sparing women at low-risk from low-value care. Ultimately, a robust prediction model would facilitate the transition from a glucocentric model-of-care to an individualised and holistic approach to this widespread public health problem.

Translating prediction models into clinical care is challenging.[45–47] Previous efforts of addressing this clinical prediction problem have been hampered by the use of methods, which increase the risk of biassed predictions limiting the transportability of developed models to new but related populations (manuscript submitted for publication, 2020). Thus, rigorous and robust methods have been adopted for model development and validation in this study. Methods have been framed by the learnings from our critical appraisal of existing models and will be guided by TRIPOD statement.[26]

Limitations
Use of routine-collected healthcare data
The development data set was created using routinely-collected healthcare data. This data was collected contemporaneously, and in a prospective fashion, however, they were not collected specifically for the purposes of this study. In prediction modelling studies, the use of routinely collected data enables the accruement of a greater number of events, which increases power to consider a greater number of candidate predictors without risking overfitting. However, the retrospective direction of enquiry creates the possibility of poor-quality data for both predictors and outcome, potential unmeasured predictors and as such careful evaluation of missing data and application of appropriate methods to address it are essential to minimise the effect on performance and applicability of developed models.[48]

Maternal death during pregnancy or any other complications that preclude delivery at the hospital will not be captured within the source perinatal outcomes database.

Varying diagnostic criteria
Diagnostic criteria used for GDM are controversial. Some professional societies endorse the criteria initially proposed by the International Association of Diabetes and Pregnancy Study Groups but disagreement persists.[4 6 49] There is also the acknowledgement that the optimal diagnostic strategy may vary depending on the characteristics of the local population.[1 8 9] The ideal prognostic prediction model would perform adequately across populations defined by a range of diagnostic criteria. Addressing this challenge will require developed models to be externally validated across these different populations.

Addressing treatment paradox regarding insulin use
Addressing the treatment paradox (in this case with insulin) is a challenge in prediction modelling studies. The traditional approach has been to accept predictions in the context of current care. However, this does not remove the possibility that a potentially useful model may appear to perform poorly due to the confounding effect of the judicious application of effective interventions to individual's whom clinicians subjectively assess to be at high risk of the outcome of interest.

Two solutions to address the problem of treatment paradox in prediction modelling studies have been advocated.[50] First, the use of treatments suspected to confound the predictor-outcome relationship can be set as a predictor in the final model. Second, the use of such effective treatments can be included within a composite outcome to be predicted. For this study, both approaches were considered but deemed inappropriate. For the former, the inclusion of the requirement for insulin therapy as a predictor is not possible as this information is not available at the intended moment of prediction—the time of GDM diagnosis, usually around 24 to 28 weeks gestation. For the later, inclusion of the requirement for insulin therapy within the composite outcome would impair its interpretability as this outcome occurs at a significantly higher frequency than the other component outcomes (31% vs approximately 10% based on our prior work).[44] This is likely to lead to a less meaningful composite that is primarily driven by the need for insulin therapy and no longer predicts what we want (adverse pregnancy outcomes).

While many promising novel approaches have been proposed in the statistical literature, such as multistate modelling or marginal structural models for 'treatment drop-ins',[51 52] at time of writing all are primarily based on empirical data and are yet to be applied to clinical prediction problems.

The three possible results from the sensitivity analysis to evaluate the effect of including the decision to treat with insulin will be informative and may be interpreted as follows. If the sensitivity analyses find that the inclusion of the decision to treat with insulin within the outcome:

1. Positively affects model performance, then this suggests the presence of treatment paradox, that is, pregnancy complications are more likely to occur in the absence of insulin therapy;
2. Has no significant effect on model performance then this suggests that the model is robust with predictive performance not affected by the decision to treat, that is, the absolute risk of adverse pregnancy outcomes for an individual woman with GDM is not affected by insulin therapy;
3. Negatively affects model performance, then this would suggest that adverse pregnancy outcomes are more likely to occur in women treated with insulin, and thus imply more 'severe' GDM or a harmful effect for this treatment. (unlikely)

The effect of treatment with insulin will be further evaluated using an IPTW algorithm to weight women according to their propensity of having been treated and transformation of the logistic model into a multinomial model. This multinomial model will have four categories depending on the occurrence of the composite pregnancy outcome and whether the women have received treatment with insulin or not.

The target population to whom the prediction model applies

The focus of this model and eventual clinical risk calculator is on those women who develop GDM and has been developed to address the priorities of frontline healthcare workers and services on the potential for risk stratified care for the one in five women who are diagnosed with GDM. Future work, should consider whether learnings from this project can be applied to a broader population, including pregnant women without GDM in particular those with maternal overweight or obesity.

ETHICS AND DISSEMINATION

This study has been approved by the Human Research Ethics Committee of Monash Health (RES-19–0000713L). This study will be conducted in accordance with the principles of the Declaration of Helsinki and the National Statement on Ethical Conduct in Human Research (2018).[53 54] All analyses will be conducted using non-identifiable data extracted from a pre-existing data set. The data is collected as part of routine clinical care for the primary purpose of improving the quality of pregnancy care. Consent was not obtained for the secondary use of this data because it is not practical to do so, and this research is consistent with the primary purpose for which it was collected. This study has been registered on the Australian and New Zealand Clinical Trials Registry (ACTRN12620000915954).[55] Results will be disseminated via presentation at scientific meetings and publication in peer-reviewed journals.

Author affiliations
[1]Monash Centre for Health Research and Implementation, School of Public Health and Preventative Medicine, Monash University, Clayton, Victoria, Australia
[2]Diabetes Unit, Monash Health, Clayton, Victoria, Australia
[3]Monash Women's Program, Monash Health, Clayton, Victoria, Australia
[4]Diabetes and Endocrinology Units, Monash Health, Clayton, Victoria, Australia
[5]CIBER Epidemiology and Public Health, Madrid, Comunidad de Madrid, Spain
[6]Clinical Biostatistics Unit, Hospital Ramon y Cajal, Madrid, Madrid, Spain
[7]Clinical Biostatistics Unit, Hospital Universitario Ramon y Cajal, Madrid, Madrid, Spain
[8]WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Shamil D. Cooray http://orcid.org/0000-0002-6825-4440
Jacqueline A. Boyle http://orcid.org//0000-0002-3616-1637
Javier Zamora http://orcid.org/0000-0003-4901-588X
Borja M. Fernández Félix http://orcid.org/0000-0002-8798-019X
John Allotey http://orcid.org/0000-0003-4134-6246
Shakila Thangaratinam http://orcid.org/0000-0002-4254-460X
Helena J. Teede http://orcid.org/0000-0001-7609-577X

REFERENCES

1 American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2019*. *Diabetes Care* 2019;42:S13–28.

2 International Diabetes Federation. *Prevalence of gestational diabetes mellitus (GDM), % Brussels, Belgium: international diabetes Federation*. 9th edn, 2019. https://diabetesatlas.org/data/en/indicators/14/

3 Buchanan TA, Xiang AH, Page KA. Gestational diabetes mellitus: risks and management during and after pregnancy. *Nat Rev Endocrinol* 2012;8:639–49.

4 Metzger BE, Gabbe SG, Persson B, *et al*. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy: response to Weinert. *Diabetes Care* 2010;33:e98–82.

5 Metzger BE, Lowe LP, HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991–2002.

6 Nankervis A, McIntyre HD, Moses RG, *et al*. ADIPS consensus guidelines for the testing and diagnosis of hyperglycaemia in pregnancy in Australia and New Zealand 2014.

7 National Institute for Health and Care Excellence. Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. *London* 2015.

8 Feig DS, Berger H, Donovan L, *et al*. Diabetes and pregnancy. *Can J Diabetes* 2018;42:S255–82.

9 Committee on Practice Bulletins-Obstetrics. ACOG practice Bulletin No. 190: gestational diabetes mellitus. *Obstet Gynecol* 2018;131:e49–64.

10 Rudland VL, Wong J, Yue DK, *et al*. Gestational diabetes: seeing both the forest and the trees. *Curr Obstet Gynecol Rep* 2012;1:198–206.

11 Scifres C, Feghali M, Althouse AD, *et al*. Adverse outcomes and potential targets for intervention in gestational diabetes and obesity. *Obstet Gynecol* 2015;126:316–25.

12 Huet J, Beucher G, Rod A, *et al*. Joint impact of gestational diabetes and obesity on perinatal outcomes. *J Gynecol Obstet Hum Reprod* 2018;47:469–76.

13 Goldstein RF, Abell SK, Ranasinha S, *et al*. Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. *JAMA* 2017;317:2207–25.

14 Yuen L, Wong VW, Simmons D. Ethnic disparities in gestational diabetes. *Curr Diab Rep* 2018;18:68.

15 Hughes AE, Nodzenski M, Beaumont RN, *et al*. Fetal genotype and maternal glucose have independent and additive effects on birth weight. *Diabetes* 2018;67:1024–9.

16 Wan CS, Abell S, Aroni R, *et al*. 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;11:809–17.

17 Cooray SD, Wijeyaratne LA, Soldatos G, *et al*. The unrealised potential for predicting pregnancy complications in women with gestational diabetes: a systematic review and critical appraisal. *Int J Environ Res Public Health* 2020;17:3048.

18 Cooray SD, Boyle JA, Soldatos G, *et al*. 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.

19 Wan CS, Nankervis A, Teede H, *et al*. Ethnicity and gestational diabetes mellitus care: providers' and patients' perspectives. *Qual Health Res* 2020.

20 Royston P, Moons KGM, Altman DG, *et al*. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.

21 Altman DG, Vergouwe Y, Royston P, *et al*. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.

22 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. Second edition. New York, London: Springer International Publishing, 2019.

23 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–31.

24 Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG: Int J Obstet Gy* 2017;124:423–32.

25 Steyerberg EW, Moons KGM, van der Windt DA, *et al*. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.

26 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.

27 Moons KGM, Altman DG, Reitsma JB, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.

28 Barnes RA, Wong T, Ross GP, *et al*. A novel validated model for the prediction of insulin therapy initiation and adverse perinatal outcomes in women with gestational diabetes mellitus. *Diabetologia* 2016;59:2331–8.

29 Egan AM, Bogdanet D, Griffin TP, *et al*. A core outcome set for studies of gestational diabetes mellitus prevention and treatment. *Diabetologia* 2020;63:1120–7.

30 Montori VM, Permanyer-Miralda G, Ferreira-González I, *et al*. Validity of composite end points in clinical trials. *BMJ* 2005;330:594–6.

31 Teede HJ, Harrison CL, Teh WT, *et al*. Gestational diabetes: development of an early risk prediction tool to facilitate opportunities for prevention. *Aust N Z J Obstet Gynaecol* 2011;51:499–504.

32 Tomlinson TM, Mostello DJ, Lim K-H, *et al*. Fetal overgrowth in pregnancies complicated by diabetes: development of a clinical prediction index. *Arch Gynecol Obstet* 2018;298:67–74.

33 Brand JS, West J, Tuffnell D, *et al*. Gestational diabetes and ultrasound-assessed fetal growth in South Asian and white European women: findings from a prospective pregnancy cohort. *BMC Med* 2018;16:203.

34 Powe CE, Allard C, Battista M-C, *et al*. Heterogeneous contribution of insulin sensitivity and secretion defects to gestational diabetes mellitus: table 1. *Diabetes Care* 2016;39:1052–5.

35 Benhalima K, Van Crombrugge P, Moyson C, *et al*. Characteristics and pregnancy outcomes across gestational diabetes mellitus subtypes based on insulin resistance. *Diabetologia* 2019;62:2118–28.

36 Black MH, Sacks DA, Xiang AH, *et al*. The relative contribution of prepregnancy overweight and obesity, gestational weight gain, and IADPSG-defined gestational diabetes mellitus to fetal overgrowth. *Diabetes Care* 2013;36:56–62.

37 Magann EF, Doherty DA, Chauhan SP, *et al*. Pregnancy, obesity, gestational weight gain, and parity as predictors of Peripartum complications. *Arch Gynecol Obstet* 2011;284:827–36.

38 Moons KGM, Wolff RF, Riley RD, *et al*. Probast: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1–33.

39 Vergouwe Y, Steyerberg EW, Eijkemans MJC, *et al*. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–83.

40 Morris TP, White IR, Carpenter JR, *et al*. Combining fractional polynomial model building with multiple imputation. *Stat Med* 2015;34:3298–317.

41 Tibshirani R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B* 1996;58:267–88.

42 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–74.

43 Egan AM, Dunne FP, Biesty LM, *et al*. Gestational diabetes prevention and treatment: a protocol for developing core outcome sets. *BMJ Open* 2019;9:e030574.

44 Abell SK, Boyle JA, Earnest A, *et al*. Impact of different glycaemic treatment targets on pregnancy outcomes in gestational diabetes. *Diabet. Med.* 2019;36:177–83.

45 Wyatt JC, Altman DG. Commentary: prognostic models: clinically useful or quickly forgotten? *BMJ* 1995;311:1539–41.

46 Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201–9.

47 Kleinrouweler CE, Cheong-See FM, Collins GS, *et al*. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol* 2016;214:e36:79–90.

48 Moons KGM, de Groot JAH, Bouwmeester W, *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 2014;11:e1001744.

49 National Institute for Health and Care Excellence. Diabetes in pregnancy: management from preconception to the postnatal period (NICE guideline [NG3]). London; 2015.

50 Cheong-See F, Allotey J, Marlin N, *et al*. Prediction models in obstetrics: understanding the treatment paradox and potential solutions to the threat it poses. *BJOG: Int J Obstet Gy* 2016;123:1060–4.

51 Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007;26:2389–430.

52 Sperrin M, Martin GP, Pate A, *et al*. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Stat Med* 2018;37:4142–54.

53 The World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 18th WMA General Assembly. Helsinki; 1964.

54 The National Health and Medical Research Council, The Australian Research Council, Universities Australia. National statement on ethical conduct in human research 2007 (updated 2018). Canberra Commonwealth of Australia; 2007.

55 Australian and New Zealand Clinical Trials Registry. The prediction modelling for risk-stratified care for women with gestational diabetes (personal GDM) study: calculating the individualised risk of adverse outcomes for women with gestational diabetes (ACTRN12620000915954) Sydney, Australia: NHMRC clinical trials centre, University of Sydney, 2020. Available: https://www.anzctr.org.au/ACTRN12620000915954.aspx [Accessed 25 Sep 2020].

56 Dobbins TA, Sullivan EA, Roberts CL, *et al*. Australian National birthweight percentiles by sex and gestational age, 1998–2007. *Med J Aust* 2012;197:291–4.

57 Australian Bureau of Statistics. 1249.0 - Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG). Canberra Commonwealth Government; 2016. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1249.0