

Development and validation of an early pregnancy risk score for the prediction of gestational diabetes mellitus in Chinese pregnant women

Si Gao,^{1,2} Junhong Leng ,³ Hongyan Liu,³ Shuo Wang,⁴ Weiqin Li,⁴ Yue Wang,³ Gang Hu ,⁵ Juliana C N Chan,^{6,7} Zhijie Yu,⁸ Hong Zhu ,^{1,2} Xilin Yang^{1,2}

To cite: Gao S, Leng J, Liu H, *et al.* Development and validation of an early pregnancy risk score for the prediction of gestational diabetes mellitus in Chinese pregnant women. *BMJ Open Diab Res Care* 2020;**8**:e000909. doi:10.1136/bmjdr-2019-000909

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2019-000909>).

SG, JL and XY contributed equally.

Received 14 September 2019
Revised 25 February 2020
Accepted 15 March 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Hong Zhu;
zhuhong@tmu.edu.cn

ABSTRACT

Objective To develop and validate a set of risk scores for the prediction of gestational diabetes mellitus (GDM) before the 15th gestational week using an established population-based prospective cohort.

Methods From October 2010 to August 2012, 19 331 eligible pregnant women were registered in the three-tiered antenatal care network in Tianjin, China, to receive their antenatal care and a two-step GDM screening. The whole dataset was randomly divided into a training dataset (for development of the risk score) and a test dataset (for validation of performance of the risk score). Logistic regression was performed to obtain coefficients of selected predictors for GDM in the training dataset. Calibration was estimated using Hosmer-Lemeshow test, while discrimination was checked using area under the receiver operating characteristic curve (AUC) in the test dataset.

Results In the training dataset (total=12 887, GDM=979 or 7.6%), two risk scores were developed, one only including predictors collected at the first antenatal care visit for early prediction of GDM, like maternal age, body mass index, height, family history of diabetes, systolic blood pressure, and alanine aminotransferase; and the other also including predictors collected during pregnancy, that is, at the time of GDM screening, like physical activity, sitting time at home, passive smoking, and weight gain, for maximum performance. In the test dataset (total=6444, GDM=506 or 7.9%), the calibrations of both risk scores were acceptable (both p for Hosmer-Lemeshow test >0.25). The AUCs of the first and second risk scores were 0.710 (95% CI: 0.680 to 0.741) and 0.712 (95% CI: 0.682 to 0.743), respectively (p for difference: 0.9273).

Conclusion Both developed risk scores had adequate performance for the prediction of GDM in Chinese pregnant women in Tianjin, China. Further validations are needed to evaluate their performance in other populations and using different methods to identify GDM cases.

INTRODUCTION

Gestational diabetes mellitus (GDM) is prevalent, affecting about 16.4% of women globally and 25.0% in the Southeast Asia region.¹ GDM is associated with both short-term and long-term adverse health consequences for both the mother and her offspring. Women

Significance of this study

What is already known about this subject?

► Gestational diabetes mellitus (GDM) is prevalent globally. Lifestyle modification before the 15th gestational week can reduce the risk of GDM.

What are the new findings?

► We have developed a set of clinical risk scores for the prediction of GDM among Chinese pregnant women before the 15th gestational week and at the screening for GDM.
► The performance of the two risk scores was adequate with good calibration and moderate discrimination.
► The first risk score (including six baseline predictors) was preferentially recommended with respect to its acceptable validation and relative simplicity.

How might these results change the focus of research or clinical practice?

► Further validations are needed to evaluate the performance of the risk scores in other populations. In addition, randomized controlled trials are required to verify whether women identified with high risk of GDM by our risk scores can benefit more from early lifestyle intervention than those identified with low risk, so that lifestyle intervention can be done in a more cost-effective manner.

with GDM are at increased risk of perinatal morbidity.² These women are also at particular high risk of diabetes and cardiovascular disease in their later life.^{3 4} Offspring born to women with GDM are at increased risk of obesity in childhood⁵ and cardiovascular disease traits in adulthood.⁶

Several randomized controlled trials (RCTs) demonstrated that lifestyle intervention among patients with GDM could improve pregnancy outcomes and reduce insulin resistance in the female offspring around 5 years of age.^{7 8} However, there were no studies reporting that intervention of GDM during pregnancy was able to reduce

long-term risk of diabetes in this high risk group of women or reduce risk of childhood obesity in offspring of GDM mothers.^{9 10} So it is critical to prevent the occurrence of GDM. In this regard, several published RCT studies, such as UPBEAT,¹¹ RADIEL,¹² St CARLOS,¹³ have demonstrated that the interventions on modifiable risk factors or lifestyle during pregnancy could decrease the incidence of GDM among pregnant women. Early interventions, for example with the Mediterranean diet, have shown benefits in women even at low risk of GDM¹³ or diagnosed as GDM.¹⁴ Furthermore, our meta-analysis¹⁵ found that lifestyle modification before the 15th gestational week (GW) could reduce the risk of GDM, but such intervention turned out ineffective once the pregnancy advanced beyond the 15th GW. Besides, we also showed that benefits of lifestyle intervention was not limited to overweight or obese women but also extended to women with normal body weight prior to pregnancy. Therefore, the key issue is to identify the group at high risk of GDM before the 15th GW or in early pregnancy so that lifestyle intervention can be done in a more cost-effective manner.

To achieve this purpose, some risk scores have been developed for the prediction of GDM.¹⁶ However, till now, almost all the risk score models have been derived from European or North American countries, such as UK,¹⁷ Germany,¹⁸ Netherlands,¹⁹ Canada,²⁰ America,^{21 22} Australia^{23 24}, and only a few were from Asian²⁵ or African population.²⁶ Thus, a specific risk model targeted to Asian population is urgently needed because of the heterogeneity of different ethnicities. Besides, the internal and external validity of some previous risk scores might be limited due to relative small sample size,^{17 22} retrospective²⁷ or cross-sectional design,²⁶ or single-center source of sample.^{19 24 25} In addition, changes of lifestyle and behaviors during pregnancy were not considered in most risk score models.¹⁶

Thus, the current study, using an established population-based prospective cohort in Tianjin, China, aimed to develop and validate risk scores for the prediction of GDM based on baseline characteristics and during-pregnancy modifiable risk factors.

MATERIALS AND METHODS

Study population and settings

This study was conducted in Tianjin, a metropolitan city in Northern China, ranking fourth in population size (14 millions in 2012) among Chinese cities. Antenatal care in Tianjin urban districts was delivered by a three-tiered antenatal care network (consisting of primary, secondary, and tertiary hospitals) in a relatively structured manner.^{28 29} In brief, all pregnant women were registered at a primary hospital and received the antenatal care until the 32nd GM. Then, they were referred to one of the secondary or tertiary care hospitals of their choice for continued care till delivery.

From October 2010 to August 2012, 22 302 pregnant women were registered to receive their antenatal care and attended the screening for GDM. The detailed methods of establishment of this cohort were described previously.^{30–32}

Screening for and diagnosis of GDM

A two-step GDM screening procedure, which was initiated in 1998, was used for the screening of GDM. First, all pregnant women were offered 50g 1-hour glucose challenge test (GCT) at primary hospitals between 24th and 28th weeks of gestation. Then, women with plasma glucose (PG) at GCT ≥ 7.8 mmol/L were referred to Tianjin Women and Children's Health Center for a standard 75g 2-hour oral glucose tolerance test (OGTT).

When The International Association of the Diabetes and Pregnancy Study Groups (IADPSG) criteria were developed in 2010, we changed the old WHO's criteria for GDM to the IADPSG's. GDM was defined by meeting any one of the cut-off values: fasting PG ≥ 5.1 mmol/L, 1-hour PG ≥ 10.0 mmol/L, or 2-hour PG ≥ 8.5 mmol/L.³³ However, to maintain the logistic and operation of the screening and management system, that is, GCT at primary care hospitals and OGTT at a centralized GDM clinic within Tianjin Women and Children's Health Care Center, we continued to use a two-step procedure to identify GDM. Considerations of use of the two-step procedure were available in previous publications.^{30–32}

Data collection

Data were collected longitudinally using self-administered questionnaires, anthropometric and laboratory measurements at two time points: at registration for pregnancy (≤ 15 th GW, mean \pm SD: 10.2 \pm 1.9) and at the time of GCT (24th and 28th GWs, mean \pm SD: 24.8 \pm 2.5).^{30–32} Firstly, baseline information, for example, demographic and socioeconomic information, lifestyle, personal and family history of disease was collected at registration; then, at the time of GCT, information on changeable lifestyle were remeasured and recorded, such as sleeping time and quality, smoking, physical activity, and weight gain from registration to GCT.

The definitions of the variables were as follows: maternal age at registration was calculated as the period in years from the date of birth to the date of registration. Family history of diabetes was defined as having one or more first degree relatives with diabetes. Active smoking before pregnancy or during pregnancy was defined as continuously smoking one or more cigarettes per day for at least 6 months before pregnancy or smoking one or more cigarettes per day during pregnancy. Passive smoking information was collected by asking "are you currently exposed to cigarette smoking from others in working and/or living places before pregnancy or during this pregnancy?" Information on sleeping status during pregnancy was collected by asking two questions: "how many hours of sleep (including nap) did you get during the index pregnancy?" and "how did you feel

about your sleep quality during the index pregnancy, good, moderate or poor?”³⁴ Physical activity (including occupational, commuting, leisure-time, and housework physical activity) during pregnancy was assessed and categorized into low level and middle-to-high level.³¹ Sitting time at home referred to the hours daily spent on sitting at home, including watching TV, reading, using the computer, and other sitting times at home, including meal time. Detailed definitions of these variables could be referred to our previous reports.^{30–32}

Maternal height, weight, waist circumference, and blood pressure (BP) were measured by uniformly trained nurses at primary care hospitals using a standardized protocol.³⁰ Body weight at registration was treated as prepregnancy weight due to small weight gain during the first 12 GWs.³⁵ Weight gain from registration to GCT was calculated as the difference in body weight from registration to GCT. Weight change during pregnancy was also assessed using gestational weight gain rate (GWGR, kg/week) according to the following formula:
$$GWGR = \frac{\text{Weight}_{\text{at the time of GCT}} - \text{Weight}_{\text{at initial registration for pregnancy}}}{\text{GW}_{\text{at the time of GCT}} - \text{GW}_{\text{at initial registration for pregnancy}}}$$

We categorized GWGR as inadequate, adequate, or excessive according to the 2009 Institute of Medicine guidelines.³⁶ Body mass index (BMI) was calculated as weight in kilogram divided by the square of body height in meter. Obesity and overweight were defined by the criteria recommended by the Working Group on Obesity in China,³⁷ that is, underweight: BMI <18.5 kg/m², normal weight: BMI 18.5 to 23.9 kg/m², overweight: BMI 24.0 to 27.9 kg/m² and obesity: BMI ≥28.0 kg/m².

ABO blood types and serum alanine aminotransferase (ALT) were measured after an overnight fasting. ABO blood types were determined by hemagglutination reactions between antigens and antibodies by the slide method. ALT was measured using an automated enzymatic method (Toshiba TBA-120FR, Japan).

Women who had any of the following conditions were excluded from data analysis: who had history of type 1 or type 2 diabetes before pregnancy, who got registered to receive their first antenatal care at >the 15th GW, or who did not complete the two-step GDM screening procedure.

Statistical analysis

IBM SPSS Statistics V.19.0 (IBM SPSS, Chicago, Illinois, USA) was used to perform the statistical analyses. The characteristics of the study population were summarized by means±SD for continuous variables and by percentages for categorical variables. Characteristics at first antenatal care visit and the change of lifestyle during pregnancy (from prepregnancy to GCT) was compared between GDM and non-GDM groups using Student’s t-test or χ^2 test where appropriate.

The dataset was randomly divided into two parts using a computer-generated random number: the training dataset and the test dataset, with the ratio of sample size of 2:1. The training dataset was used to develop the risk score and the test dataset was used to validate

its performance. In view of the relatively big sample size, simple randomization method, rather than block randomization or stratified randomization, was used for allocation without need to consider any variables or characteristics of participants.

Risk score development

We chose to develop two sets of risk scores, one only including predictors collected at the first antenatal care visit for early prediction of GDM and the other also including risk factors collected during pregnancy, that is, at the time of GCT.

In the training dataset, binary logistic regression was performed to obtain ORs and 95% CIs of related factors for GDM. The dependent variable was the development of GDM, and the candidate independent variables were the characteristics of participants before and during pregnancy which had a univariate significance level of $p < 0.20$ and/or were judged to be of clinical importance and/or had been proved to be associated with GDM by our previous analyses. In multivariate logistic regression, enter method rather than stepwise method was used for the selection of independent variables to avoid overfitting, and only statistically significant variables ($p < 0.05$) remained in the model. The shrinkage factor was calculated using $(\chi^2 - k) / \chi^2$, where χ^2 denotes the likelihood ratio χ^2 and k the number of the predictors in the model (below 0.85 raises concern of overfitting). If necessary, the regression coefficients of the predictors were multiplied by the shrinkage factor (uniformly shrunken) to adjust for optimism.³⁸ All continuous independent variables, such as age, BMI, weight gain, systolic BP, ALT, and body height were included in the model without being categorized with the aim of minimizing the loss of information caused by dichotomization.

The interactions between independent variables were assessed by generating new variables (the value of “1” represented any of two variables was abnormal, and the value of “0” represented both of two variables was normal) and recruiting them into the model.

Validation of the developed risk scores

Validation of the developed risk score was performed in the test dataset. Calibration and discrimination were used to check the performance of the developed risk score. First, Hosmer-Lemeshow χ^2 test was used to check the calibration. Pregnant women in the test dataset were divided into deciles according to their predicted probability of GDM. The observed and expected probabilities of GDM in the deciles were compared using the Hosmer-Lemeshow test (df=8). A p value of more than 0.10 indicated similarity in the predicted and observed probability or an acceptable calibration. Second, discrimination was assessed using area under the receiver operating characteristic curve (AUC). Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) at different cut-off points of the risk score

was calculated for possible use of the risk score at other different antenatal care scenarios.

RESULTS

Characteristics of the study population

Among 22 302 pregnant women, we sequentially excluded 21 women who had history of type 1 or type 2 diabetes before pregnancy, 936 who registered and attended their first antenatal care in more than the 15th GW, 1163 women who did not undergo GCT, and 851 women who had a positive GCT but did not undergo OGTT. Finally, 19 331 women were included in the analysis (online supplementary figure 1).

Among 19 331 eligible participants, 1485 women (7.68%) developed GDM. Women with GDM were more likely to be older, Han-ethnicity, multiparous, have non-AB blood type, have habitual use of tobacco before pregnancy, and have family history of diabetes in first-degree relatives. They also had higher level of BMI, waist circumference, BP, and ALT, but shorter height than those without GDM. Besides, during pregnancy, passive smoking, shorter (<7 hours/day) or longer (≥ 9 hours/day) duration of sleep, and more sitting time at home was also more common among GDM cases than their counterparts. No significant differences were found between two groups with respect to education, weight gain, and physical activity during pregnancy (table 1).

In addition, we compared the basic characteristics of participants between the training dataset and the test dataset and found the two groups had good similarity with respect to almost all profiles, except for active smoking during pregnancy ($p=0.006$) and sitting hours per day during pregnancy ($p=0.029$), which demonstrated that our simple random allocation method was reasonable (data not shown).

Risk score development

The training dataset had 979 or 7.6% GDM cases ($n=12887$). The selected predictors, their regression coefficients (β), and ORs for the first risk score and the second one were listed, respectively, in model 1 and model 2 of table 2.

Among the potential predictors collected at the first antenatal care visit, non-AB blood type, active smoking before pregnancy, additive interactions between overweight and high ALT, and additive interactions between overweight and height were no longer significant in multivariate analysis and thus not included in the first risk score. Waist circumference was also not recruited in the final model because of its collinearity with BMI at registration. Consequently, the first risk score consisted of six baseline predictors: maternal age, BMI at registration, body height, SBP, ALT, and family history of diabetes in first-degree relatives. Their β coefficients were shown in model 1 of table 2.

Based on the first risk score, we further tested the predictive values of the variables collected at GCT, like

physical activity during pregnancy, sitting time at home during pregnancy, weight gain from registration to GCT, passive smoking during pregnancy, sleeping quality and sleeping time during pregnancy, as well as the additive interaction between overweight and passive smoking. We found that sleeping quality and sleep time during pregnancy and additive interaction between overweight and passive smoking were not significant in multivariate analysis. Consequently, the second risk score consisted of physical activity during pregnancy, sitting time at home during pregnancy, weight gain from registration to GCT, passive smoking during pregnancy, as well as the predictors in the first risk score. Their β coefficients were shown in model 2 of table 2.

The shrinkage factor of model 1 and model 2 was 0.985 and 0.968, respectively, significantly higher than the value of overfitting criteria ($\lambda=0.85$), indicating that the performance of the two models was only overestimated by 1.5% and 3.2%, respectively. Thus, it was not vitally necessary for our data to make adjustment of parameters by shrinkage factor. Based on the unadjusted values of β coefficients, the final risk score of GDM in model 1 and model 2 were constructed as follows:

Model 1: GDM risk score= $0.0941 \times$ maternal age (year)+ $0.1278 \times$ BMI at registration (kg/m^2)+ $0.0093 \times$ SBP (mm Hg)+ $0.6816 \times$ Log $_{10}$ (ALT) (U/L)+ $0.5129 \times$ family history of diabetes (1 if yes, 0 if no)- $0.0270 \times$ body height (cm)-5.7469.

Model 2: GDM risk score= $0.0978 \times$ maternal age (year)+ $0.1366 \times$ BMI at registration (kg/m^2)+ $0.013 \times$ SBP (mm Hg)+ $0.7004 \times$ Log $_{10}$ (ALT) (U/L)+ $0.4909 \times$ family history of diabetes (1 if yes, 0 if no)- $0.0215 \times$ body height (cm)- $0.2374 \times$ physical activity during pregnancy (1 if middle-to-high level, 0 if low level)+ $0.1825 \times$ sitting time at home (1 if <2 hours/day, 2 if 2–4 hours/day, 3 if >4 hours/day)+ $0.0351 \times$ weight gain (kg)+ $0.3058 \times$ passive smoking during pregnancy (1 if yes, 0 if no)-8.0732.

Risk scores validation

The test dataset had 506 or 7.9% GDM cases ($n=6444$). The first risk score (model 1) had an acceptable calibration, with the predicted probabilities of GDM being similar to the observed probabilities (χ^2 for Hosmer-Lemeshow test= 10.052 , $p>0.25$). In the same way, the second risk score (model 2) had similar predicted probabilities of GDM with the observed ones (χ^2 for Hosmer-Lemeshow test= 7.995 , $p>0.25$) (figure 1).

The first risk score achieved an AUC of 0.710 (95% CI: 0.680 to 0.741). After further including physical activity during pregnancy, sitting time at home during pregnancy, weight gain from registration to GCT, and passive smoking during pregnancy, the discrimination of the second risk score slightly improved (AUC: 0.712, 95% CI: 0.682 to 0.743), but not statistically significant ($p=0.9273$) (figure 2).

In the second risk score, we also evaluated the predictive value of GWGR by recruiting GWGR, instead of weight gain from registration to GCT, into the second risk score.

Table 1 Characteristics of participants according to the occurrence of GDM

Characteristic	Non-GDM (n=17846)	GDM (n=1485)	T/ χ^2	P value
Variables at first antenatal care visit				
Maternal age, years	28.4±2.9	29.6±3.2	-13.31	<0.001
Han ethnicity	17019 (95.4)	1435 (96.6)	5.082	0.024
Education ≤12 years	2966 (16.6)	265 (17.9)	1.528	0.216
Parity ≥1	617 (3.5)	69 (4.6)	5.661	0.017
Active smoking prepregnancy	799 (4.8)	93 (7.2)	14.319	<0.001
Passive smoking prepregnancy	11985 (72.3)	958 (74.1)	2.063	0.151
Family history of diabetes	1340 (8.2)	205 (16.0)	91.163	<0.001
Non-AB blood type	15876 (89.6)	1345 (91.3)	4.467	0.035
Systolic blood pressure, mm Hg	105.3±10.6	108.5±11.0	-11.126	<0.001
Diastolic blood pressure, mm Hg	68.2±7.7	70.5±8.0	10.848	<0.001
ALT,U/L*	1.22±0.21	1.29±0.23	-10.753	<0.001
Height, cm	163.2±4.7	162.7±4.8	4.172	<0.001
Weight, kg	59.0±9.5	64.0±11.4	-16.529	<0.001
Waist circumference, cm	78.7±8.6	82.9±9.7	-15.125	<0.001
BMI at registration, kg/m ²	22.1±3.3	24.2±3.9	-19.382	<0.001
BMI at registration group, kg/m ²				
<18.5	1847 (10.4)	53 (3.6)	393.422	<0.001
18.5–	11673 (65.4)	748 (50.4)		
24.0–	3291 (18.4)	471 (31.7)		
28.0–	1029 (5.8)	213 (14.3)		
Variables at GCT				
Active smoking during pregnancy	47 (0.4)	1 (0.1)	1.235	0.266
Passive smoking during pregnancy	8438 (65.0)	710 (69.7)	9.303	0.002
Weight gain from registration to GCT, kg	7.5±3.5	7.4±3.6	0.345	0.73
Sleeping time during pregnancy, hours/day				
<7	239 (1.9)	22 (2.2)	6.66	0.036
7–	5414 (43.0)	385 (38.8)		
9–	6935 (55.1)	584 (58.9)		
Sleeping quality during pregnancy				
Good	4989 (38.4)	355 (34.8)	5.849	0.054
Middle	7708 (59.4)	639 (62.6)		
Bad	284 (2.2)	27 (2.6)		
Physical activity during pregnancy				
Low	2091 (17.5)	184 (19.6)	2.846	0.092
Moderate to high	9887 (82.5)	753 (80.4)		
Sitting time at home during pregnancy, hours/day				
<2	1315 (10.4)	68 (6.8)	14.053	0.001
2–4	5934 (46.7)	472 (47.1)		
>4	5451 (42.9)	463 (46.2)		
Plasma glucose at GCT, mmol/L	6.3±1.3	9.4±1.5	-76.822	<0.001
Fasting glucose at OGTT, mmol/L	4.5±0.3	5.2±0.7	-30.225	<0.001
1-hour glucose at OGTT, mmol/L	8.1±1.2	10.2±1.6	-41.607	<0.001
2-hour glucose at OGTT, mmol/L	6.8±1.0	8.6±1.7	-36.415	<0.001
HbA1c% at OGTT, %	4.9±0.4	5.2±0.6	-14.205	<0.001

Data were reported in mean±SD or number (%), which were calculated based on the non-missing data.

*ALT was log-transformed

ALT, alanine aminotransferase; BMI, body mass index; GCT, 50 g 1-hour glucose challenge test; GDM, gestational diabetes mellitus; HbA1c, hemoglobin A1c; OGTT, 75 g 2-hour oral glucose tolerance test.

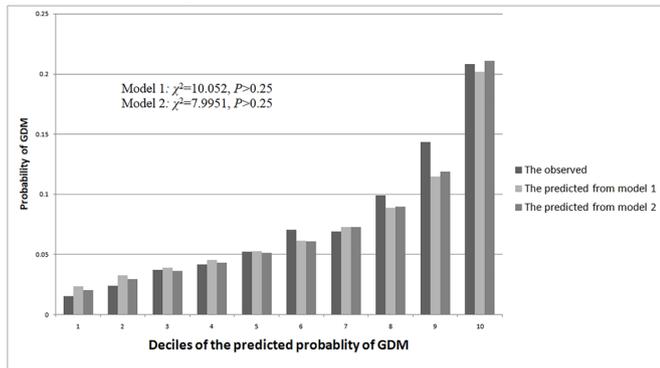


Figure 1 The predicted and observed probability of gestational diabetes mellitus (GDM) based on model 1 and model 2 in the test dataset.

We found that the proportion of inadequate, adequate, or excessive weight gain in our whole population were 17.8%, 24.9%, and 53.7%, respectively. In univariate analysis, excessive GWG might increase the risk of GDM relative to adequate GWG, with the OR (95% CI) of 1.26 (1.07 to 1.48) ($p=0.006$), but this significance disappeared in the multivariate analysis (model 2), with the OR (95% CI) of 0.95 (0.76 to 1.18) ($p=0.641$) (data not shown). Besides, recruiting GWGR in the model 2 did not improve the AUC of model 2 (0.706, 95% CI: 0.684 to 0.729) when compared with including weight gain into the model (AUC: 0.712, 95% CI: 0.682 to 0.743). Based on all these results, and considering the simplicity of the model, we still recommended using weight gain, rather than GWGR, as a GDM predictor in the second risk score.

Sensitivity, specificity, PPV, and NPV of the risk scores at different cut-off points were summarized in table 3. To avoid missed diagnosis of GDM (false negatives), the relatively lower cut-off values of risk score should be recommended. For example, at the cut-off point of 2.50 for the first risk score, sensitivity, specificity, PPV, and NPV were 93.3%, 25.1%, 9.6%, and 97.8%, respectively. For the

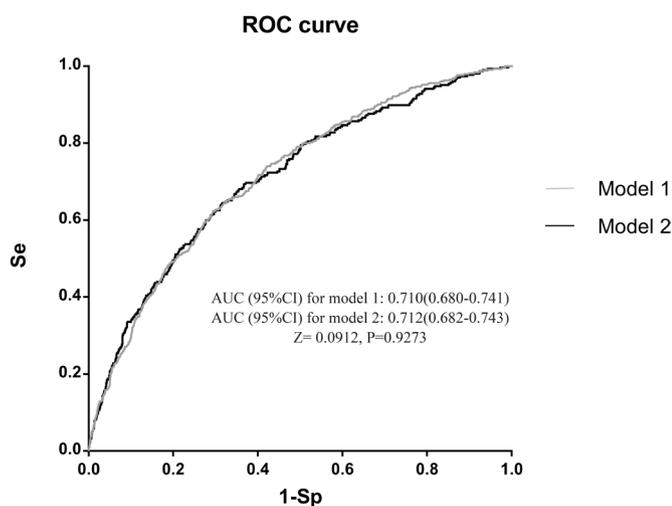


Figure 2 Area under receiver operating characteristic (ROC) curve (AUC) of model 1 and model 2 in test dataset. Se, sensitivity; Sp, specificity.

second risk score, at the cut-off point of 4.7, sensitivity, specificity, PPV, and NPV were 93.5%, 20.9%, 9.2%, and 97.4%, respectively. Using these two cut-off values, more than 93% patients with GDM could be identified, with the missed diagnosis rate of less than 7%. If we applied a risk score of 2.80 in model 1 as a threshold to identify women “at high risk” for GDM, 57.4% of all women would need to undergo OGTT or receive preventive intervention. The corresponding PPV and NPV was 11.2% and 96.7%, respectively. For the second risk score, if the threshold to proceed to diagnostic test was set at, for example, 5.1, 54.7% of all women would be subjected to OGTT. The PPV and NPV were 11.1% and 96.6%, respectively.

Moreover, sensitivity analysis was conducted to think about the OGTT missing data (1163 who did not undergo GCT, and 851 women who had a positive GCT but not undergo OGTT). First, basic characteristics (variable in table 1) of this population were compared with those who had completed the two-step GDM screening procedure and were recruited in our analysis ($n=19931$). Almost all profiles were similar between two groups, except for education and parity. Women who did not receive GCT or OGTT were more likely to be multiparous (7.5% vs 3.5%) or have education less than 12years (23.5% vs 16.7%) than their counterparts. Second, multiple imputation was conducted,³⁹ and the missing OGTT measurements were estimated on the basis of the results of the GCT tests as well as the characteristics of the participants (variables in table 1). We found that the included predictors in the two risk score models based on the imputed data were same as those based on the data without imputation, and no significant changes on calibration and discrimination were observed. Besides, excluding 686 women who had at least once parity did not influence the performance of the two models (data not shown).

DISCUSSION

Our study developed and validated a set of early pregnancy risk scores for the prediction of GDM using a representative sample of 19331 pregnant women in Tianjin, China. We found that six predictors collected at the first antenatal care visit (maternal age, BMI, height, systolic BP, ALT, and family history of diabetes in first-degree relatives) and four during-pregnancy modifiable risk factors (physical activity, sitting time at home, passive smoking, and weight gain from registration to GCT) were associated with an increasing risk of GDM. The first risk score including only baseline variables and the second risk score including both baseline and during pregnancy variables had similar and acceptable calibration (both p for Lemeshow test >0.25) and discrimination (AUC for the first and the second risk score was 0.710 (95% CI: 0.680 to 0.741) and 0.712 (95% CI: 0.682 to 0.743), respectively).

Although numerous risk factors for GDM have been identified, the ability to accurately identify women before or early in pregnancy who are at the high risk of GDM and

Table 3 Sensitivity, specificity, predictive values at selected risk score and the predicted probability of GDM in test dataset

Risk score cut-off point	Model 1						Model 2					
	Sensitivity %	Specificity %	PPV* %	NPV* %	Predicted probability of GDM %	Risk score cut-off point	Sensitivity %	Specificity %	PPV* %	NPV* %	Predicted probability of GDM %	
1.30	100.0	0.1	7.9	100.0	1.1465	3.50	100.0	0.2	7.9	100.0	1.0219	
1.80	99.5	1.8	7.9	97.7	1.8949	4.00	99.7	2.1	8.0	98.8	1.6738	
2.30	96.6	14.3	8.8	98.0	3.0861	4.60	95.1	16.5	8.8	97.5	3.0084	
2.40	95.6	19.1	9.1	98.1	3.3997	4.70	93.5	20.9	9.2	97.4	3.3144	
2.45	94.7	21.8	9.4	98.0	3.5678	4.75	90.6	23.8	9.2	96.7	3.4784	
2.50	93.3	25.1	9.6	97.8	3.7438	4.80	89.9	26.3	9.4	96.8	3.6502	
2.60	90.1	31.2	10.0	97.4	4.1214	4.90	87.9	32.4	10.0	96.9	4.0187	
2.65	88.5	34.3	10.3	97.2	4.3235	5.00	85.7	38.1	10.6	96.9	4.4226	
2.70	86.4	37.8	10.6	97.0	4.5351	5.04	84.0	40.8	10.8	96.8	4.5948	
2.75	84.8	41.2	10.9	97.0	4.7566	5.07	82.7	42.5	10.9	96.6	4.7282	
2.80†	82.1	44.8	11.2	96.7	4.9883	5.10†	81.8	44.4	11.1	96.6	4.8651	
2.90	78.2	51.3	12.0	96.5	5.4842	5.20	77.5	50.6	11.8	96.3	5.3494	
3.00	74.0	57.7	13.0	96.3	6.0262	5.30	72.3	56.3	12.4	96.0	5.8790	
3.10	66.9	63.2	13.4	95.7	6.6180	5.40	69.7	62.1	13.5	96.0	6.4573	
3.20	63.2	68.8	14.7	95.6	7.2635	5.50	64.8	67.3	14.4	95.7	7.0883	
3.30	56.8	74.0	15.7	95.3	7.9666	5.60	59.9	72.1	15.5	95.5	7.7758	
3.40	51.3	78.0	16.6	94.9	8.7312	5.70	53.7	76.6	16.4	95.1	8.5239	
3.60	40.5	85.6	19.3	94.4	10.4621	6.00	39.7	86.5	20.0	94.4	11.1729	
3.80	28.3	90.5	20.2	94.6	12.4892	6.34	25.6	93.0	23.8	93.6	15.0179	
4.35	12.4	97.6	30.6	92.9	19.8308	6.70	12.7	96.9	25.9	92.9	20.2103	

*PPV and NPV were calculated based on the prevalence of GDM in testing dataset (ie, 506/6444)

†The suggested cut-off points.

GDM, gestational diabetes mellitus; NPV, negative predictive value; PPV, positive predictive value.

could benefit most from interventions remains limited. Only a few studies have summarized their results and developed predictions model or scoring systems to estimate the risk of GDM individually. Recently, Lamain-de Ruiter *et al*¹⁶ performed an external validation of 12 published GDM prediction models. He found that most of the published models showed acceptable discrimination and calibration, with the AUCs ranged from 0.67 to 0.78. Of these 12 models, 2 models^{19 24} were assessed by another researcher in a cohort of 510 Finland women.⁴⁰ However, the results showed that both models underestimated the GDM incidence in this population. These inconsistent results suggested the marked heterogeneity of GDM in different populations.

Our risk scores based on Chinese population achieved similar calibration and discrimination as those based on European or North American populations. Some similar predictors^{30 40–45 16} has been identified in our analysis, like maternal age, maternal BMI, family history of diabetes, systolic BP, and ALT level. However, our study did not observe significant association of GDM with history of GDM and ethnicity. The non-significant association of GDM with GDM history was partly due to the overwhelming proportion of nulliparous (95.9%) women in our cohort who had no previous pregnancy and no chance to get GDM. Hence, based on our data, the role of GDM history could not be assessed thoroughly. As for ethnicity, the ethnic heterogeneity of our study was lower than that of Lamain-de Ruiter's report which included Caucasian, African, Asian, mixed, and other ethnicity. This might be partly the reason for our insignificant association between ethnicity and GDM risk. External validation is needed to test whether our risk score could be generalized to other Asian populations, as well as to non-Asian ethnicities.

Short body height was observed to be associated with an increased risk of GDM in our data. Similar conclusion was also drawn from a Korean study.⁴¹ Some scholars⁴⁶ suspected that the association between short height and increased risk of GDM was particularly seen among Asians and may not warrant biological plausibility for use as a GDM predictor. In our opinion, even though body height might not have causal association with GDM, the significant improvement of the performance of risk score after inclusion of height in the model convinced us that keeping this variable in the model was reasonable. The role of height in predicting GDM among non-Asian populations should be further studied.

Moreover, four modifiable risk factors during pregnancy (physical activity, sitting time at home, passive smoking, and weight gain from registration to GCT) were found to be associated with an increased risk of GDM. Although adding these four indicators into the risk score model did not increase the AUC significantly, their value for clinical intervention was potentially huge. Lifestyle intervention such as increasing physical activity, decreasing sitting time, keeping reasonable weight gain, and avoiding passive smoking should be promoted

during pregnancy. Nevertheless, till now, it seems still unclear whether women at high risk of GDM compared with those at low risk could benefit more from early intervention. Further studies are needed to clarify which kind of intervention strategy (whole-population strategy or high-risk population strategy) is more cost-effective.

In our study, two risk scores and their corresponding cut-off values were developed. To simplify the utility of risk scores in clinical practice, the first model was preferentially recommended with respect to its acceptable validation and relative simplicity (only including six easily detected variables). The cut-off value of 2.80 or 3.00 could be used before the 15th GW to identify the high-risk pregnant women. However, these recommended threshold values were arbitrary. To determine the optimal threshold applied to diagnostic testing, more information should be obtained, such as the feasibility of the model in practice, the preferences of obstetricians, the incidence of GDM, and the costs and the availability of diagnostic testing.¹⁹

The strength of our study included that the risk score models was developed and evaluated based on a prospective cohort with a large sample size and enough GDM cases. An unselected population of pregnant women registered in the three-tiered antenatal care network could guarantee good representativeness of our sample.

However, there were still some limitations in our study. First, our risk score was derived and validated by pregnant women population with GDM identified using a two-step procedure. Further validation studies in other care settings such as use of different antenatal care system and different GDM identification procedures are warranted. Presumably, it is needed to upcalibrate the absolute risk of GDM in those places where one-step GDM identification procedures are in use. Second, diet information before and during pregnancy were not collected when taking account of the feasibility of the survey and the simplicity of the models. Instead, BMI at registration and gestational weight gain was selected as predictors based on the present evidence that these two anthropometrics were closely associated with die quality.^{47 48} To our knowledge, only one risk score model had included diet as a predictor of GDM.²³ Further studies are needed to explore the sensitive and valid diet-related items for GDM prediction. Moreover, external validation of our risk scores is required to evaluate the generalizability and applicability of our findings in other populations and different settings.

In conclusion, we have developed a set of clinical risk scores for the prediction of GDM among pregnant women before the 15th GW and at the time of GCT. The performance of the two risk scores was adequate with good calibration and moderate discrimination. Further validation is needed to evaluate the performance of the risk scores in other populations. In addition, RCTs are urgently required to verify whether GDM high-risk women identified by our models can benefit more from early lifestyle intervention than the low-risk ones, so

that lifestyle intervention can be done in a more cost-effective manner.

Author affiliations

- ¹Department of Epidemiology and Biostatistics, School of Public Health, Tianjin Medical University, Tianjin, China
²Tianjin Key Laboratory of Environment, Nutrition and Public Health, Tianjin Medical University, Tianjin, China
³Department of Child Health, Tianjin Women and Children's Health Center, Tianjin, China
⁴Project Office, Tianjin Women and Children's Health Center, Tianjin, China
⁵Chronic Disease Epidemiology Laboratory, Pennington Biomedical Research Center, Baton Rouge, Louisiana, USA
⁶Department of Medicine and Therapeutics, Prince of Wales Hospital-International Diabetes Federation Centre of Education, The Chinese University of Hong Kong, Hong Kong, China
⁷Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China
⁸Population Cancer Research Program and Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada

Acknowledgements We thank all doctors, nurses and research staffs at 65 primary care hospitals, 6 district-level women and children's health centers (WCHCs), and Tianjin Women and Children's Health Center (TWCHC) for their participation in this study.

Contributors XY, HL, GH, JCNC and ZY conceived and designed the study; all authors, except GH, ZY, and JCNC contributed to the collection of the data. SG and JL analyzed the data and wrote the first draft; XY revised the draft critically for important intellectual content. All authors gave critical comments and contributed to the writing of the manuscript; SG, JL, XY and HZ take full responsibility for the work as a whole, including the study design, access to data, and the decision to submit and publish the manuscript.

Funding This work was supported by National Key Research and Development Program of China (Grant nos: 2018YFC1313900 and 2018YFC1313903), and National Natural Science Foundation of China (Grant nos: 81870549, 81602922, and 81900724).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study ethics was approved by the Ethics Committee for Clinical Research of TWCHC. Written informed consent was obtained from all pregnant women before data collection.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The raw data were generated at TWCHC. Data supporting the findings of this study are available from the corresponding author on reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which 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

Junhong Leng <http://orcid.org/0000-0002-6742-0830>
 Gang Hu <http://orcid.org/0000-0002-6172-8017>
 Hong Zhu <http://orcid.org/0000-0003-0501-1211>

REFERENCES

- Guariguata L, Linnenkamp U, Beagley J, *et al*. Global estimates of the prevalence of hyperglycaemia in pregnancy. *Diabetes Res Clin Pract* 2014;103:176–85.
- HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, *et al*. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008;358:1991–2002.
- Song C, Lyu Y, Li C, *et al*. Long-Term risk of diabetes in women at varying durations after gestational diabetes: a systematic review and meta-analysis with more than 2 million women. *Obes Rev* 2018;19:421–9.
- Li J, Song C, Li C, *et al*. Increased risk of cardiovascular disease in women with prior gestational diabetes: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2018;140:324–38.
- Hillier TA, Pedula KL, Schmidt MM, *et al*. Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. *Diabetes Care* 2007;30:2287–92.
- Kaseva N, Vääräsmäki M, Sundvall J, *et al*. Gestational diabetes but not prepregnancy overweight predicts for cardiometabolic markers in offspring twenty years later. *J Clin Endocrinol Metab* 2019;104:2785–95.
- Crowther CA, Hillier JE, Moss JR, *et al*. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* 2005;352:2477–86.
- Landon MB, Spong CY, Thom E, *et al*. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* 2009;361:1339–48.
- Gillman MW, Oakey H, Baghurst PA, *et al*. Effect of treatment of gestational diabetes mellitus on obesity in the next generation. *Diabetes Care* 2010;33:964–8.
- Landon MB, Rice MM, Varner MW, *et al*. Mild gestational diabetes mellitus and long-term child health. *Diabetes Care* 2015;38:445–52.
- Poston L, Bell R, Croker H, *et al*. Effect of a behavioural intervention in obese pregnant women (the upbeat study): a multicentre, randomised controlled trial. *Lancet Diabetes Endocrinol* 2015;3:767–77.
- Koivusalo SB, Rönö K, Klemetti MM, *et al*. Gestational diabetes mellitus can be prevented by lifestyle intervention: the Finnish gestational diabetes prevention study (RADIEL): a randomized controlled trial. *Diabetes Care* 2016;39:24–30.
- Assaf-Balut C, García de la Torre N, Durán A, *et al*. A Mediterranean diet with additional extra virgin olive oil and pistachios reduces the incidence of gestational diabetes mellitus (GDM): a randomized controlled trial: the St. Carlos GDM prevention study. *PLoS One* 2017;12:e0185873.
- Assaf-Balut C, García de la Torre N, Durán A, *et al*. Medical nutrition therapy for gestational diabetes mellitus based on Mediterranean diet principles: a subanalysis of the ST Carlos GDM prevention study. *BMJ Open Diabetes Res Care* 2018;6:e000550.
- Song C, Li J, Leng J, *et al*. Lifestyle intervention can reduce the risk of gestational diabetes: a meta-analysis of randomized controlled trials. *Obes Rev* 2016;17:960–9.
- Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, *et al*. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ* 2016;354:i4338.
- Nanda S, Savvidou M, Syngelaki A, *et al*. Prediction of gestational diabetes mellitus by maternal factors and biomarkers at 11 to 13 weeks. *Prenat Diagn* 2011;31:135–41.
- Gabbay-Benziv R, Doyle LE, Blitzer M, *et al*. First trimester prediction of maternal glycemic status. *J Perinat Med* 2015;43:283–9.
- van Leeuwen M, Opmeer BC, Zweers EJK, *et al*. Estimating the risk of gestational diabetes mellitus: a clinical prediction model based on patient characteristics and medical history. *BJOG* 2010;117:69–75.
- Thériault S, Giguère Y, Massé J, *et al*. Early prediction of gestational diabetes: a practical model combining clinical and biochemical markers. *Clin Chem Lab Med* 2016;54:509–18.
- Donovan BM, Breheny PJ, Robinson JG, *et al*. Development and validation of a clinical model for preconception and early pregnancy risk prediction of gestational diabetes mellitus in nulliparous women. *PLoS One* 2019;14:e0215173.
- Badon SE, Zhu Y, Sridhar SB, *et al*. A Pre-Pregnancy biomarker risk score improves prediction of future gestational diabetes. *J Endocr Soc* 2018;2:1158–69.
- Schoenaker DAJM, Vergouwe Y, Soedamah-Muthu SS, *et al*. Preconception risk of gestational diabetes: development of a prediction model in nulliparous Australian women. *Diabetes Res Clin Pract* 2018;146:48–57.
- 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.
- Phaloprakarn C, Tangjitgamol S, Manusirivithaya S. A risk score for selective screening for gestational diabetes mellitus. *Eur J Obstet Gynecol Reprod Biol* 2009;145:71–5.
- Nombo AP, Mwanri AW, Brouwer-Brolsma EM, *et al*. Gestational diabetes mellitus risk score: a practical tool to predict gestational diabetes mellitus risk in Tanzania. *Diabetes Res Clin Pract* 2018;145:130–7.
- Sweeting AN, Wong J, Appelblom H, *et al*. A first trimester prediction model for gestational diabetes utilizing aneuploidy and

- pre-eclampsia screening markers. *J Matern Fetal Neonatal Med* 2018;31:2122–30.
- 28 Yang X, Hsu-Hage B, Zhang H, *et al*. Gestational diabetes mellitus in women of single gravidity in Tianjin City, China. *Diabetes Care* 2002;25:847–51.
 - 29 Yang X, Tian H, Zhang F, *et al*. A randomised translational trial of lifestyle intervention using a 3-tier shared care approach on pregnancy outcomes in Chinese women with gestational diabetes mellitus but without diabetes. *J Transl Med* 2014;12:290.
 - 30 Leng J, Shao P, Zhang C, *et al*. Prevalence of gestational diabetes mellitus and its risk factors in Chinese pregnant women: a prospective population-based study in Tianjin, China. *PLoS One* 2015;10:e0121029.
 - 31 Leng J, Liu G, Zhang C, *et al*. Physical activity, sedentary behaviors and risk of gestational diabetes mellitus: a population-based cross-sectional study in Tianjin, China. *Eur J Endocrinol* 2016;174:763–73.
 - 32 Leng J, Wang P, Shao P, *et al*. Passive smoking increased risk of gestational diabetes mellitus independently and synergistically with prepregnancy obesity in Tianjin, China. *Diabetes Metab Res Rev* 2017;33:dmr.2861.
 - 33 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.
 - 34 Wang H, Leng J, Li W, *et al*. Sleep duration and quality, and risk of gestational diabetes mellitus in pregnant Chinese women. *Diabet Med* 2017;34:44–50.
 - 35 Fattah C, Farah N, Barry SC, *et al*. Maternal weight and body composition in the first trimester of pregnancy. *Acta Obstet Gynecol Scand* 2010;89:952–5.
 - 36 Km R, Al Y. *Weight gain during pregnancy: reexamining the guidelines*. Washington, 2009.
 - 37 Chen C, Lu FC, Department of Disease Control Ministry of Health, PR China. The guidelines for prevention and control of overweight and obesity in Chinese adults. *Biomed Environ Sci* 2004;17 Suppl:1–36.
 - 38 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
 - 39 van der Heijden GJMG, Donders ART, Stijnen T, *et al*. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.
 - 40 Huvinen E, Eriksson JG, Stach-Lempinen B, *et al*. Heterogeneity of gestational diabetes (GDM) and challenges in developing a GDM risk score. *Acta Diabetol* 2018;55:1251–9.
 - 41 Jang HC, Min HK, Lee HK, *et al*. Short stature in Korean women: a contribution to the multifactorial predisposition to gestational diabetes mellitus. *Diabetologia* 1998;41:778–83.
 - 42 Fraser A, Harris R, Sattar N, *et al*. Alanine aminotransferase, gamma-glutamyltransferase, and incident diabetes: the British women's heart and health study and meta-analysis. *Diabetes Care* 2009;32:741–50.
 - 43 Hedderon MM, Ferrara A. High blood pressure before and during early pregnancy is associated with an increased risk of gestational diabetes mellitus. *Diabetes Care* 2008;31:2362–7.
 - 44 Li J, Wang P, Zhang C, *et al*. Short body height and Pre-pregnancy overweight for increased risk of gestational diabetes mellitus: a population-based cohort study. *Front Endocrinol* 2018;9:00349.
 - 45 Leng J, Zhang C, Wang P, *et al*. Plasma levels of alanine aminotransferase in the first trimester identify high risk Chinese women for gestational diabetes. *Sci Rep* 2016;6:27291.
 - 46 Babu GR, Nakamura A, Jurišić Eržen D. Commentary: short body height and Pre-pregnancy overweight for increased risk of gestational diabetes mellitus: a population-based cohort study. *Front Endocrinol* 2018;9:00575.
 - 47 Parker HW, Tovar A, McCurdy K, *et al*. Associations between pre-pregnancy BMI, gestational weight gain, and prenatal diet quality in a national sample. *PLoS One* 2019;14:e0224034.
 - 48 Ancira-Moreno M, Vadillo-Ortega F, Rivera-Dommarco Juan Ángel, *et al*. Gestational weight gain trajectories over pregnancy and their association with maternal diet quality: results from the PRINCESA cohort. *Nutrition* 2019;65:158–66.