Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

DEVELOPMENT AND VALIDATION OF A NON-INVASIVE RISK PREDICTION MODEL FOR EARLY IDENTIFICATION OF GESTATIONAL DIABETES MELLITUS AMONG PREGNANT WOMEN IN PAKISTAN

Dr. Haider Kamal[*1], Dr. Muhammad Khan[2], Huma Shafique[3], Dr. Tayyaba Mehmood[4], Dr. Muhammad Shoaib[5], Dr. Iqra Hafeez[6], Khubib Mahmood[7], Jawad Zada[8]

[*1]*MBBS, Post Graduate*
[2]*Medical Officer*
[3]*Dietitian, MPhil Human Nutrition and Dietetics*
[4]*FCPS-I in Obstetrics and Gynaecology*
[5]*MBBS (KE), BSc (PU)*
[6]*Medical Officer*
[7]*Medix College of Health Management Science and Institute of Nursing, Chakdara, Pakistan*
[8]*Medix College of Health Management Science and Institute of Nursing, Chakdara, Pakistan*

[*1]haiderkamal541@gmail.com, [2]drmkniazi@gmail.com, [3]humashafiq195@gmail.com, [4]tayyabamehmood122@gmail.com, [5]sheikhmshoaib37@gmail.com, [6]iqrahafeez333@gmail.com

DOI: https://doi.org/10.5281/zenodo.2390755

Keywords
Gestational diabetes mellitus; risk prediction model; risk score; non-invasive; oral glucose tolerance test; antenatal screening; Pakistan

Article History
Received: 01 January 2026
Accepted: 17 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Dr. Haider Kamal

Abstract

Background. *Gestational diabetes mellitus (GDM) is becoming more common in low- and middle-income countries, including Pakistan. A simple, affordable and easy-to-use risk prediction model for early identification of high-risk women is therefore needed. This study set out to develop and validate a non-invasive risk prediction model, the Non-INvaSive Pregnancy RIsk ScoRE (INSPIRE), which uses risk factors already reported in the literature to identify Pakistani pregnant women at high risk of GDM.*

Methods. *In a cross-sectional study, 550 pregnant women attending antenatal clinics at one tertiary and two secondary hospitals in islamabad were enrolled between weeks 29 and 33 of pregnancy. The data were randomly split into a derivation set (422 women, 80%) and a validation set (94 women, 20%). Researchers interviewed the women about sociodemographic details and family history of diabetes, measured mid-upper arm circumference (MUAC), and took obstetric history and oral glucose tolerance test (OGTT) results from hospital records. Multivariable logistic regression on the derivation set produced the predictor weights. Calibration was tested with Pearson's chi-square goodness-of-fit test, and discrimination was measured with the area under the curve (AUC), both in the validation set.*

Results. *GDM affected 32 % of the women (136 of 550). The final score uses six predictors: maternal age, MUAC, family history of diabetes, previous GDM, a previous bad obstetric outcome, and previous macrosomia. In the validation set the score was well calibrated (chi-square 29.55, p = 0.08) and showed acceptable discrimination (AUC 0.721, 95% CI 0.61 to 0.83), with 74.1% sensitivity and 59.4% specificity.*

Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Conclusion. *INSPIRE is a non-invasive risk prediction model that allows early identification of Pakistani pregnant women at high risk of GDM and separates them from those at low risk, which can spare many women an unnecessary OGTT.*

INTRODUCTION

Over recent decades GDM prevalence in Pakistan has climbed steeply, from 6.3% in 2003 to 19% in 2018. GDM means abnormal glucose tolerance that begins or is first noticed during pregnancy. It raises the risk of short-term problems such as caesarean delivery, pregnancy-induced hypertension, premature rupture of membranes, and bleeding before and after birth. It also has long-term consequences: roughly 17% to 33% of South Asian women with GDM develop type 2 diabetes within 5 to 10 years of that pregnancy. Identifying high-risk women early is therefore important so that treatment can begin in time.

The current diagnostic approach is a one-step 75 g OGTT performed between weeks 24 and 28. The woman drinks a 75 g glucose load, and blood glucose is checked at 1 hour, 2 hours and often 3 hours. GDM is diagnosed if one or more readings reach or exceed the set thresholds. The test is burdensome in several ways:

- Many women find it unpleasant because of nausea, vomiting and repeated blood draws, which leads to non-attendance.
- It takes about 4 to 5 hours, forcing daily-wage workers to lose a day's pay and homemakers to arrange household cover.
- The cost is a major barrier for low- and middle-income families and for those living in remote areas.

Because of these obstacles, a simple and inexpensive risk score would be valuable. A risk score estimates the probability that a person has, or will develop, a condition by combining several risk factors, and it should be validated in the population where it will be used.

Many countries already use risk scores to offer the OGTT only to higher-risk women, but existing scores fit South Asian women poorly because of differences in ethnicity, social background, body composition and obstetric factors. Some are also hard to use in low-resource settings. A Chinese score, for instance, requires alanine transaminase (ALT), which is not a routine blood test in pregnancy in many such settings. A Singaporean score relies on an artificial-intelligence model and trained staff to enter data. A locally suited score would help direct OGTT referrals to those who need them and spare low-risk women. This study therefore aimed to develop and validate INSPIRE, a non-invasive risk prediction model for early identification of GDM among Pakistani pregnant women, using literature-based risk factors and checking it against the 2-hour 75 g OGTT.

Methods

Design and setting

This was a cross-sectional study. Validity was assessed against a gold standard, the 2-hour 75 g OGTT. Data were gathered from February to May 2025 at the pims Hospital

Recruitment and data collection

Women attending antenatal clinics between weeks 29 and 33 of pregnancy were approached and chosen purposively. To be included they needed a single baby, an age of 18 to 45 years, and a documented OGTT done between weeks 24 and 28. Women were not invited if they had known diabetes, heart disease or kidney failure, were taking medicines that affect glucose metabolism, or had incomplete records. After the study was explained, written consent was obtained.

Trained data collectors did three things:

1. Reviewed medical records for obstetric history (previous GDM, abortion, miscarriage, stillbirth, intrauterine death, macrosomia) and OGTT results.
2. Interviewed women face to face with a structured questionnaire on age, education, occupation, language, household income and family history of diabetes.
3. Measured height, weight and MUAC.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Sample size

Using PASS software, the calculation called for 402 women. It assumed an expected GDM prevalence of 19%, 80% power, a 5% two-sided significance level, and a precision of 0.10 around an AUC of 0.80.

Statistical analysis

The 550 enrolled women were randomly divided into a derivation set of 404 (80%) and a validation set of 96 (20%). Normally distributed continuous variables were summarised as mean and standard deviation, skewed ones as median and interquartile range, and categories as counts and percentages. Groups were compared with t-tests or chi-square tests as appropriate. Analysis was done in Stata version 17.

Building INSPIRE

Candidate predictors came from an extensive literature review. They covered maternal age, education, occupation, household income, MUAC, family history of diabetes, parity, previous GDM, adverse obstetric history (miscarriage, abortion, stillbirth, intrauterine death) and previous macrosomia (a baby over 4 kg).

First, each factor was tested alone against GDM using univariate logistic regression, keeping those with p below 0.25. Age was split into under 25 and 25 or older, since Asian women from 25 years upward are at higher risk. MUAC above 32 cm was treated as obesity, since it corresponds to a pre-pregnancy BMI above 30 kg/m². Next, forward stepwise multivariable logistic regression was used, retaining variables that were statistically significant (p below 0.05) or clinically important. Each regression coefficient was rounded to the nearest whole number to give that factor's score.

Validating INSPIRE

In the validation set, women were grouped into ten equal bands (deciles) by predicted GDM probability, and observed rates were compared with expected rates. Pearson's chi-square test judged calibration, with p above 0.05 counted as acceptable. Discrimination was judged by the AUC, from a receiver operating characteristic (ROC) curve, which plots sensitivity against the false-positive rate (1 minus specificity). Sensitivity, specificity, positive and negative predictive values and the Youden index were computed at many cut-offs so that a practical threshold for antenatal clinics could be chosen.

Results

Who took part

Of the 500 women, 130 (26%) had GDM. Compared with women without GDM, those with GDM were older (average 30.3 vs 27.4 years), heavier (77.2 vs 69.0 kg) and had a larger MUAC (31.0 vs 29.4 cm). They were far more likely to have a family history of diabetes (82.3% vs 44.9%), a prior GDM history, a prior bad obstetric outcome and prior macrosomia. Half of GDM cases (50.8%) were aged 30 or over, versus 32.7% of non-GDM women. A MUAC above 32 cm was found in 32.3% of GDM women versus 15.4% of the others. Education level, occupation, language, household income (median 50,000 PKR in both groups), height and number of pregnancies did not differ significantly. The derivation and validation sets were comparable on all characteristics except household income (p = 0.03), suggesting the random split worked well.

Building the score

The derivation set had 103 GDM cases (25.5%). Education, occupation, income and parity lost significance and were dropped. The final model contained six predictors, shown here with adjusted odds ratios (OR) and their points in the simplified score:

- **Age 25 or older:** OR 2.36 (95% CI 1.03 to 5.41), 1 point.
- **MUAC above 32 cm:** OR 2.02 (1.07 to 3.82), 1 point.
- **Family history of diabetes:** OR 3.99 (2.09 to 7.60), 1 point.
- **Previous GDM:** OR 27.27 (8.80 to 84.47), 3 points.
- **Previous bad obstetric outcome (stillbirth, abortion, miscarriage or intrauterine death):** OR 7.78 (1.88 to 32.18), 2 points.
- **Previous macrosomia (baby over 4 kg):** OR 4.35 (1.19 to 15.96), 2 points.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

The model as a whole was significant (chi-square 143.86). The total score ranges from 0 to 10 by adding the points for each "yes" answer.

Validation using the probability equation

The validation set had 27 GDM cases (28.1%). The equation can be explained in words as follows. Start with a baseline value of minus 3.41. Add 0.86 if the woman is 25 or older, 0.70 if her MUAC is above 32 cm, 1.38 if she has a family history of diabetes, 3.31 if she had GDM before, 2.05 if she had a bad obstetric outcome, and 1.47 if she previously had a baby over 4 kg. Call the total z. The probability of GDM is then e raised to z, divided by one plus e raised to z.

Calibration was good (chi-square 29.55, p = 0.08). Discrimination was acceptable, with an AUC of 0.721 (95% CI 0.61 to 0.83). The chosen probability cut-off of 0.208 gave sensitivity 74.1%, specificity 59.4%, positive predictive value 41.7%, negative predictive value 85.4% and a Youden index of 0.34.

Validation using the simplified points score

The points-based version performed almost as well, with an AUC of 0.703 (95% CI 0.59 to 0.82). A cut-off of 2 points or more was chosen, giving sensitivity 74.1%, specificity 56.5%, positive predictive value 40.0%, negative predictive value 84.8% and a Youden index of 0.31.

How the score is used

The clinician asks six yes/no questions (age band and MUAC band are each an either/or choice). Only one answer is chosen per question, and every question must be answered. The points are added up:

- Aged 25 or older: 1 point (under 25: 0).
- MUAC above 32 cm: 1 point (32 cm or less: 0).
- Family history of diabetes: 1 point.
- GDM in an earlier pregnancy: 3 points.
- Earlier stillbirth, abortion, miscarriage or intrauterine death: 2 points.
- Earlier baby weighing more than 4 kg: 2 points.

A total of 2 or more is classed as high risk and prompts referral for an OGTT.

Trade-offs at different cut-offs (in words)

The simplified score shows the usual trade-off between catching cases and avoiding false alarms:

- **1 point or more:** very sensitive (92.6%) but only 23.2% specificity, so most women would be referred.
- **2 points or more (chosen):** 74.1% sensitivity, 56.5% specificity.
- **3 points or more:** sensitivity falls to 40.7% while specificity rises to 81.2%.
- **4 points or more:** sensitivity 29.6%, specificity 91.3%.
- **5 points and above:** specificity reaches 98.5% to 100%, but sensitivity is only 29.6% and drops to 3.7% by 8 points, so most true cases would be missed.

The probability version follows the same pattern. At a very low cut-off of 0.072 it catches 92.6% of cases but with only 23.2% specificity. The chosen 0.208 cut-off sits at 74.1% sensitivity and 59.4% specificity. Beyond about 0.8, every flagged woman had GDM (100% specificity and positive predictive value), yet only about a quarter or fewer of true cases were detected.

Description of the Figures

Figure 1: predicted versus observed GDM probability by decile (validation set). This is a paired bar chart with ten risk groups from lowest to highest. In each group, one bar shows the observed share of women who had GDM and the other the share predicted by the model. In most groups the two bars are almost the same height. Risk climbs steadily across groups: about 3% in the lowest, roughly 7 to 8% in the second, about 15 to 25% in the middle groups, around 32 to 38% in the eighth, about 64 to 67% in the ninth, and roughly 91 to 92% in the highest. The one clear mismatch is the fourth group, where the observed rate (about 25%) is noticeably higher than predicted (about 11%). Overall, the model tracks reality well.

Figure 2: ROC curves (validation set). Two curves are drawn, one for the probability-based model (AUC 0.721) and one for the simplified points score (AUC 0.703). They lie almost on top of each

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

other. Both climb quickly to roughly 30% sensitivity with almost no false positives, then rise more gradually along a gentle diagonal-like path, reaching about 74% sensitivity when the false-positive rate is roughly 0.4 to 0.45, and about 92% when it is roughly 0.75, before reaching the top corner. Both stay well above the chance line, showing moderate discriminating ability, with the simplified score losing very little accuracy.

Discussion

GDM keeps rising among South Asian women, including those in Pakistan. A score that is easy to apply and performs adequately could be a first screening step, sending only high-risk women for the OGTT and easing the financial and logistical load, especially where resources are scarce. This study found six predictors and produced a score with good calibration ($p = 0.08$) and adequate discrimination (AUC 0.721, sensitivity 74.1%, specificity 59.4%).

Other scores exist but have limitations here:

- **Naylor and colleagues** built the first clinical GDM scoring system using age, race and pre-pregnancy BMI, but it was derived from Europeans, Americans and Asians living in Canada, so it may not fit Pakistani women. Still, because higher BMI and age are known risk factors, both were included in INSPIRE. Since pre-pregnancy weight is often unavailable locally, MUAC, which is easy to measure at antenatal visits, was used as a stand-in for BMI and proved a significant predictor.

- **Gao and colleagues** (China) used early-pregnancy factors including maternal age, BMI, height, systolic blood pressure, ALT and family history of diabetes, plus four modifiable factors during pregnancy. Their score achieved an AUC of 0.71 with acceptable calibration, but ALT is not a routine test in local antenatal care, the score only works for women who attend care regularly enough to record weight gain, and its many continuous variables demand skilled calculation. INSPIRE avoids this by making all six factors simple yes/no or two-band items.

- **The Tanzanian score** used only three factors (MUAC, previous stillbirth and family history of diabetes) and had a lower AUC of 0.64. It omitted maternal age and previous GDM, both of which are well-supported risk factors that proved significant in INSPIRE.

Previous GDM was a very strong predictor, and the study included a good mix of first-time and repeat mothers, helping to fill a gap in earlier scores that overlooked this factor. The link between earlier macrosomia and GDM is well established: raised maternal blood glucose crosses the placenta, and the fetus lays down extra body fat.

With the cut-off of 2 points, about half of women would be sent for an OGTT, and more than 74% of women who truly have GDM would be found, so under 26% would be missed. A high-sensitivity threshold suits a screening tool, since the priority is to avoid missing cases. Populations differ in ethnicity, body composition and obstetric factors, so local scores are needed; INSPIRE's performance matches or exceeds that of scores built elsewhere.

Strengths. To the authors' knowledge, this is the first GDM risk score for Pakistani women. It rests on an extensive literature search, uses dichotomised predictors that suit routine antenatal care, and was checked against the gold-standard 2-hour 75 g OGTT. Obstetric history came from medical records, which lowers recall bias.

Limitations.

- The hospital setting produced a higher GDM prevalence, which may introduce admission bias. However, the three hospitals serve women from very low to high socioeconomic backgrounds, supporting generalisability.
- Confidence intervals were wide for previous GDM, bad obstetric outcome and macrosomia because these were uncommon in women without GDM, though this reflects genuine differences between the groups.
- The score was not externally validated because of limited time and resources.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

- Modifiable risk factors were not collected, though the authors consider the six chosen factors sufficient.
- Women with incomplete records were excluded, which may cause selection bias, so results should be read with care.

Conclusion

The authors developed and validated INSPIRE, a non-invasive, easy-to-use risk prediction model that identifies pregnant women at high risk of GDM during routine antenatal care and so reduces unnecessary OGTTs in low-risk women. Because it relies on established risk factors, it could also help identify high-risk women as early as the first trimester, allowing targeted intervention that might prevent GDM and its short- and long-term complications. Further research should test it on an external dataset so that it can eventually be used earlier in pregnancy in community settings, for example by female health workers who visit homes in low-resource areas of Pakistan.

REFERENCES

Sheikh A, Sheikh L. Changing prevalence of gestational diabetes mellitus during pregnancy over more than a decade. J Pak Med Assoc 2020;70:1477–8.

Ramos-Leví AM, Pérez-Ferre N, Fernández MD, et al. Risk factors for gestational diabetes mellitus in a large population of women living in Spain: implications for preventative strategies. Int J Endocrinol 2012;2012:312529.

Bener A, Saleh NM, Al-Hamaq A. Prevalence of gestational diabetes and associated maternal and neonatal complications in a fast-developing community: global comparisons. Int J Womens Health 2011;3:367–73.

Muche AA, Olayemi OO, Gete YK. Effects of gestational diabetes mellitus on risk of adverse maternal outcomes: a prospective cohort study in Northwest Ethiopia. BMC Pregnancy Childbirth 2020;20:73.

Gadve SS, Chavanda S, Mukherjee AD, et al. Risk of developing type 2 diabetes mellitus in South Asian women with history of gestational diabetes mellitus: A systematic review and meta-analysis. Indian J Endocrinol Metab 2021;25:176–81.

Panel IC. International Association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010;33:676–82.

Force UPST. Screening for gestational diabetes: US preventive services task force recommendation statement. JAMA 2021;326:531–8.

Lachmann EH, Fox RA, Dennison RA, et al. Barriers to completing oral glucose tolerance testing in women at risk of gestational diabetes. Diabet Med 2020;37:1482–9.

Handlos LN, Witte DR, Almdal TP, et al. Risk scores for diabetes and impaired Glycaemia in the Middle East and North Africa. Diabet Med 2013;30:443–51.

Lindstram J, Tuomilehto J. The diabetes risk score A practical tool to predict type 2 diabetes risk. Diabetes Care 2003;26:725–31.

Harrison CL, Lombard CB, East C, et al. Risk stratification in early pregnancy for women at increased risk of gestational diabetes. Diabetes Res Clin Pract 2015;107:61–8.

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 Diabetes Res Care 2020;8:e000909.

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.

Caliskan E, Kayikcioglu F, Oztürk N, et al. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. Acta Obstet Gynecol Scand 2004;83:524–30.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Iqbal R, Rafique G, Badruddin S, et al. Increased body fat percentage and physical inactivity are independent predictors of gestational diabetes mellitus in South Asian women. Eur J Clin Nutr 2007;61:736–42.

Anand SS, Yusuf S, Vuksan V, et al. Differences in risk factors, Atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the study of health assessment and risk in ethnic groups (SHARE). Lancet 2000;356:279–84.

Kumar M, Chen L, Tan K, et al. Population-centric risk prediction modeling for gestational diabetes mellitus: A machine learning approach. Diabetes Res Clin Pract 2022;185:S0168-8227(22)00049-3.

Khan R, Ali K, Khan Z. Sociodemographic risk factors of gestational diabetes mellitus. Pak J Med Sci 2013;29:843–6.

Ahkter J, Qureshi R, Rahim F, et al. Diabetes in pregnancy in Pakistani women: prevalence and complications in an indigenous South Asian community. Diabet Med 1996;13:189–91.

Teh WT, Teede HJ, Paul E, et al. Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. Aust N Z J Obstet Gynaecol 2011;51:26–30.

Lee KW, Ching SM, Ramachandran V, et al. Prevalence and risk factors of gestational diabetes mellitus in Asia: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2018;18:494.