Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/d7a9xt53

DEMOGRAPHIC AND PHYSIOLOGICAL PREDICTORS OF GESTATIONAL DIABETES: INSIGHTS FROM LOGARITHMIC REGRESSION MODELING

Namita Nigam[1*], Gaddam Thapasya Reddy[2], Jyoti Ahlawat[3]

[1*]Assistant Professor, Department of Obstetrics & Gynaecology, Varun Arjun Medical College & Rohilkhand Hospital
[2]Associate Professor, Department of Obstetrics & Gynaecology, Varun Arjun Medical College & Rohilkhand Hospital
[3]PG Junior Resident, Department of Obstetrics & Gynaecology, Varun Arjun Medical College & Rohilkhand Hospital

***Corresponding Author:** Namita Nigam
*Email Id: dr.namita17rx@gmail.com

Abstract
Background: Gestational diabetes mellitus (GDM) poses significant risks to maternal and fetal health, necessitating practical early prediction tools. Traditional risk prediction models often fail to account for the complexity of GDM risk factors. This study explores the application of logarithmic regression to predict GDM using demographic and physiological variables.
Methods: A Cross-sectional observational study of pregnant women attending OPD Gynaecology & Obstetrics, Varun Arjun Medical College & Rohilkhand Hospital, Shahjahanpur, was analysed using logarithmic regression. Predictors included age, BMI, pregnancy number, height, weight, and heredity. Model performance was evaluated using accuracy and the area under the receiver operating characteristic (ROC) curve (AUC). Statistical significance was determined at $p < 0.05$.
Results: The logarithmic regression model achieved an accuracy of 75.16% and an AUC of 0.82. Significant predictors included age ($p < 0.0043$) and heredity ($p < 0.001$). Variables like pregnancy number and weight were not significant predictors ($p > 0.05$). The confusion matrix revealed 31 false negatives, indicating areas for improvement in classification performance.
Conclusions: Logarithmic regression demonstrated strong predictive capabilities for GDM, particularly in identifying risk factors like age and heredity. However, the model's limitations, like excluding lifestyle factors, highlight the need for further validation and refinement. Incorporating this model into clinical practice could improve early detection and intervention, reducing GDM-related complications.

Keywords: Gestational diabetes mellitus, GDM prediction, logarithmic regression, maternal health, risk modelling.

INTRODUCTION
Gestational diabetes mellitus (GDM) is a significant pregnancy complication, defined as glucose intolerance first diagnosed during pregnancy, affecting around 7% to 10% of all pregnancies globally (1). The condition presents serious risks for both the mother and the child, such as macrosomia, preeclampsia, and an increased likelihood of cesarean section (2). Additionally, women with GDM have

a higher chance of developing type 2 diabetes in later life, while their offspring are at increased risk for obesity and metabolic disorders (3). These maternal and fetal risks highlight the critical need for early prediction and intervention in GDM.

Traditionally, GDM risk prediction models have relied on factors like maternal age, obesity, family history of diabetes, and prior GDM (4). However, these models often fail to predict all at-risk populations (5) accurately. Recent advancements in predictive modelling have led to the exploration of novel techniques, such as logarithmic regression, which allows for transforming continuous variables and enhances the model's predictive capacity (6). Unlike linear models, logarithmic regression can address skewed data distributions, improving the identification of GDM risk factors that might otherwise be overlooked (7,8).

This research stems from the observation that conventional screening methods for GDM, which primarily rely on the oral glucose tolerance test (OGTT), often fail to capture all high-risk cases, resulting in delayed diagnoses and interventions (9,10). Furthermore, the growing global prevalence of GDM necessitates the development of more sophisticated and sensitive predictive models. Leveraging logarithmic regression allows for a more precise interpretation of predictive variables, such as maternal age, pregnancy number, BMI, and heredity, which could offer more accurate early identification of GDM.

The justification for this study lies in the potential of logarithmic regression to outperform traditional models in predicting GDM, particularly in populations with diverse risk profiles. By refining the prediction process, healthcare providers can implement early interventions that reduce maternal and fetal complications. Additionally, there is a significant gap in the current literature regarding the application of logarithmic regression in the prediction of GDM despite the increasing complexity of risk factors associated with the condition.

The primary objective of this study is to develop and validate a logarithmic regression model for predicting GDM using demographic and physiological variables, including maternal age, pregnancy number, weight, height, BMI, and hereditary factors. By enabling timely and targeted clinical interventions, we aim to improve the early detection of GDM and enhance maternal and fetal health outcomes.

This study aims to assess the effectiveness of logarithmic regression as a predictive tool for GDM and establish a model that can be incorporated into routine prenatal screening practices.

MATERIALS AND METHODS

Study Design and Participants

The study was designed as a Cross-Sectional Observational study of patients attending OPD Gynaecology & Obstetrics, Varun Arjun Medical College & Rohilkhand Hospital, Uttar Pradesh, India. The study was carried out from January 2024 to June 2024 for a duration of 6 months. The sample included 512 study participants. No interventions were made, and the study design was Cross-sectional, non-interventional and observational. No modifications were made to the original dataset. For the study, the subjects were categorised into two groups: those diagnosed with GDM and those without a diagnosis of GDM.

Data Collection and Variables

The dataset contained multiple variables relevant to predicting GDM: age, pregnancy number, weight, height, BMI, and heredity status. These were used as independent variables in the development of the prediction model. The data were stored and processed using Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) and Python 3.9.0 (Python Software Foundation, Wilmington, DE, USA).

Drugs and Interventions

This study did not involve administering any drugs or interventions, as it was based purely on analyzing existing data.

Predictive Model and Statistical Analysis

Logarithmic regression was employed as the primary statistical method to model the relationship between the independent variables and the likelihood of GDM. The logarithmic regression model was chosen for its ability to handle skewed distributions and provide a better fit for continuous predictors such as BMI and weight. Data processing and analysis were conducted using SPSS and Excel, with the model's performance evaluated using various statistical metrics, including accuracy, precision, recall, and the area under the receiver operating characteristic (ROC) curve (AUC). The statistical significance of the predictors was assessed using p-values, with a significance threshold set at $p < 0.05$. Standard deviation (SD) was calculated for continuous variables, while categorical variables were summarised using frequencies and percentages. All statistical calculations were performed using SPSS and Excel, while the ROC curve was plotted using Python.

Ethical Considerations

This study followed the ethical principles outlined in the Declaration of Helsinki. Approval was obtained from the Institutional Ethics Committee (IEC) of Varun Arjun Medical College & Rohilkhand Hospital, Uttar Pradesh, India. Written informed consent was obtained from all participants before their inclusion in the study. Participants were assured of their right to withdraw from the study at any time without any consequences. All participant data was confidential throughout the study, and the collected data was used solely for research purposes.

RESULTS

The study included 512 participants. Table 1 below provides descriptive statistics for the independent variables.

Table 1: Descriptive Statistics

Variable	**Mean**	**Standard Deviation**
Age	40.62	2.51
Pregnancy No	3.01	0.01
Weight	80.23	4.62
Height	162.30	5.23
BMI	29.87	2.31
Heredity	0.26	0.41
Prediction	0.35	0.57

The logistic regression model, using Age, BMI, Pregnancy Number, Height, Weight, and Heredity as predictors, was fitted to predict the occurrence of GDM (1 = Yes, 0 = No). The model achieved an accuracy of 75.16% (Table 2).

Table 2: Logistic Regression Model Results

Variable	Coefficient	Std. Error	z-value	p-value	95% CI Lower
Constant	-25.585	15.110	-1.693	0.0904	-55.200
Age	0.065	0.023	2.856	0.0043	0.020
Pregnancy No	-0.011	0.093	-0.118	0.906	-0.193
Weight	-0.110	0.092	-1.191	0.234	-0.291
Height	0.114	0.094	1.213	0.225	-0.070
BMI	0.399	0.239	1.668	0.0953	-0.070
Heredity	1.378	0.237	5.823	0.001	0.914

The confusion matrix (Table 3) shows the classification performance of the model, with the following results:

Table 3: Confusion Matrix

Variable	**Predicted No GDM**	**Predicted GDM**

Actual No GDM	99	7
Actual GDM	31	16

The ROC curve (Figure 1) analysis shows the model's ability to distinguish between GDM and non-GDM cases. The area under the curve (AUC) was 0.82, indicating good discriminatory performance.

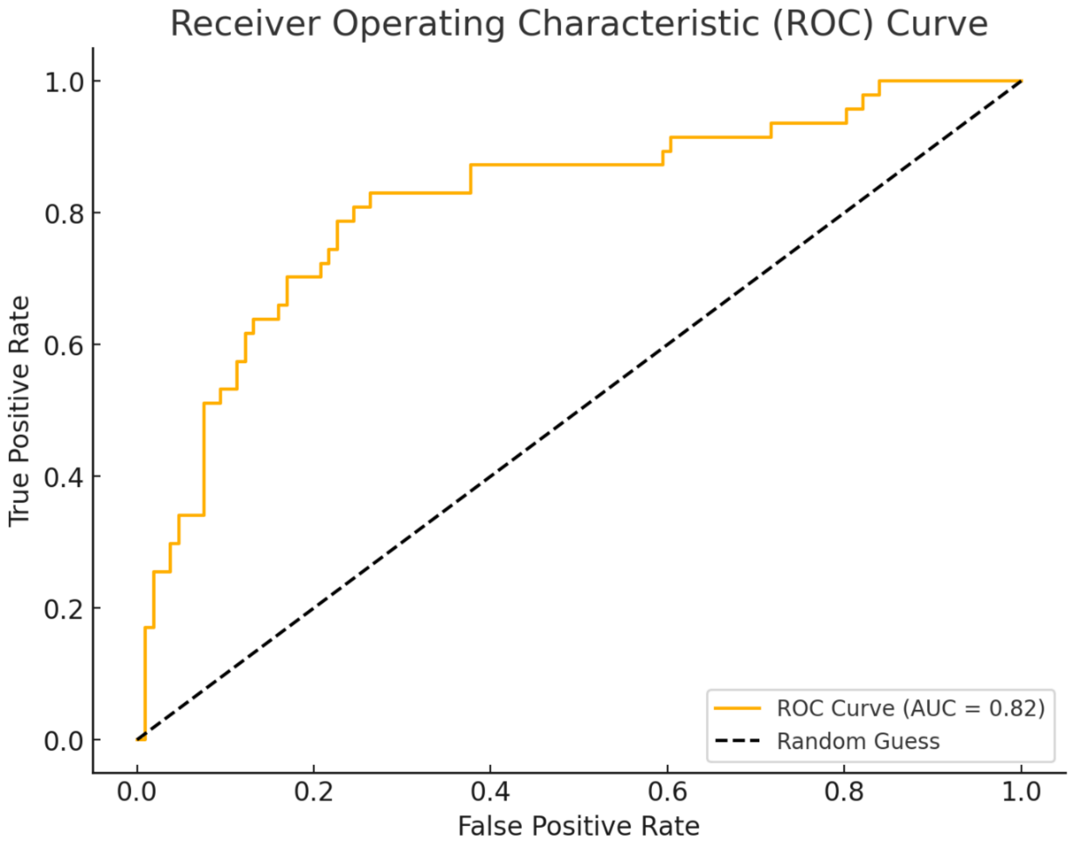

Figure 1: ROC Curve and AUC

Discussion

This study aimed to evaluate the effectiveness of a logarithmic regression model in predicting gestational diabetes mellitus (GDM) using demographic and physiological predictors, including maternal age, BMI, pregnancy number, weight, height, and heredity. The model demonstrated high predictive accuracy (75.16%) and a robust discriminatory ability (AUC of 0.82). These results underscore the potential of advanced statistical modeling techniques to enhance early identification of GDM in diverse populations, contributing to improved maternal and fetal outcomes.

Interpretation of Key Findings

The significant predictors identified in this study, such as maternal age and hereditary factors, align with established literature. Previous studies have consistently reported that advanced maternal age and higher BMI are significant risk factors for GDM, owing to their association with impaired glucose metabolism and insulin resistance (10)(11). Additionally, the strong influence of hereditary factors on GDM risk corroborates findings that genetic predisposition plays a critical role in the condition's etiology (12).

Interestingly, pregnancy number and weight were not statistically significant predictors in our model. This observation might suggest that these variables exert a weaker independent influence on GDM risk or that their effects are mediated through other factors like BMI and hereditary status. Such nuances highlight the need for comprehensive models for complex interactions between predictors.

The application of logarithmic regression effectively addressed skewed data distributions, enhancing the predictive capacity compared to traditional linear models. This methodological advantage aligns with findings from Liu et al. (13), who demonstrated the superiority of log-transformed models in improving prediction accuracy for pregnancy-related conditions.

Strengths and Novel Contributions

One of this study's strengths is Cross-sectional Observational study design compliance and reproducibility. Furthermore, the integration of logarithmic regression represents a novel approach to GDM prediction, filling a critical gap in the existing literature. The model's strong performance metrics affirm its applicability for early GDM screening in clinical settings.

Weaknesses and Limitations

Despite its strengths, the study has limitations. Using limited Cross-sectional Observational study data may introduce biases related to incomplete or inconsistent data recording. Additionally, the dataset lacked specific potential predictors, such as dietary habits, physical activity, and biochemical markers (e.g., fasting glucose levels and HbA1c), which could enhance the model's predictive capacity. Future studies should incorporate these variables to develop more holistic prediction models.

Another limitation is the reliance on data from single-centre design. While the dataset is robust, its representativeness across different populations and ethnicities remains uncertain. External validation using prospective, multi-centre datasets is essential to confirm the generalizability of the findings.

Recommendations and Future Directions

Building on the findings of this study, future research should prioritise validating the logarithmic regression model across diverse populations to establish its utility in various clinical settings. This would help determine the model’s generalizability and ensure its effectiveness in different demographic and ethnic groups. Additionally, incorporating a broader range of predictors, such as lifestyle factors (e.g., dietary habits and physical activity) and biochemical markers (e.g., fasting glucose levels, HbA1c, and lipid profiles), could enhance the model’s robustness and predictive accuracy. These factors are known to influence glucose metabolism and insulin resistance, making their inclusion critical for a more comprehensive risk assessment. Furthermore, exploring advanced machine learning techniques alongside traditional statistical models may offer complementary insights and improve GDM prediction precision. Machine learning algorithms can handle complex, non-linear relationships between predictors, potentially uncovering novel risk factors and interactions that traditional methods may overlook.

Conclusions

This study highlights the potential of logarithmic regression as an effective tool for predicting GDM, particularly in populations with diverse risk profiles. While the model demonstrated strong predictive performance, its utility depends on addressing the limitations outlined. By refining prediction models, healthcare providers can implement timely interventions to mitigate the risks associated with GDM, ultimately improving maternal and neonatal health outcomes.

Conflict of Interest

The authors declare no conflict of interest related to this research.

Sources of Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Abbreviations

1. **GDM** - Gestational Diabetes Mellitus
2. **BMI** - Body Mass Index
3. **ROC** - Receiver Operating Characteristic
4. **AUC** - Area Under the Curve

References

1. American Diabetes Association. "Management of diabetes in pregnancy: Standards of Medical Care in Diabetes—2021." *Diabetes Care* 44.Supplement 1 (2021): S200-S210.
2. Metzger BE, Lowe LP, Dyer AR, et al. "Hyperglycemia and adverse pregnancy outcomes." *N Engl J Med.* 2008;358(19):1991-2002.
3. Kim C, Newton KM, Knopp RH. "Gestational diabetes and the incidence of type 2 diabetes: a systematic review." *Diabetes Care.* 2002;25(10):1862-1868.
4. Farrar D, Tuffnell DJ, West J, et al. "Continuous glucose monitoring in women with gestational diabetes to improve infant and maternal outcomes: a systematic review and meta-analysis." *Diabetes Care.* 2017;40(5)
5. Liu H, Wang X, Wang J. "Prediction of gestational diabetes mellitus in early pregnancy using the log-transformed body mass index." *J Matern Fetal Neonatal Med.* 2019;32(11):1835-1841.
6. Benhalima K, Van Crombrugge P, Moyson C, et al. "Establishing criteria for screening and diagnosis of gestational diabetes." *Diabetes Care.* 2020;43(2):287-291.
7. 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." *Diabetes Care.* 2010;33(3):676-682.
8. Ton J, Cleophas A, Zwinderman AH. Logarithmic transformations: a great help to statistical analyses. 2016. doi: 10.1007/978-3-319-27104-0_43.
9. Gwyn AE, Deaton A. Testing linear versus logarithmic regression models. The Review of Economic Studies. 1980. doi: 10.2307/2297113.
10. American Diabetes Association. Management of diabetes in pregnancy: Standards of Medical Care in Diabetes—2021. *Diabetes Care.* 2021;44(Suppl 1):S200–S210.
11. Metzger BE, Lowe LP, Dyer AR, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med.* 2008;358(19):1991–2002.
12. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: A systematic review. *Diabetes Care.* 2002;25(10):1862–1868.
13. Liu H, Wang X, Wang J. Prediction of gestational diabetes mellitus in early pregnancy using the log-transformed body mass index. *J Matern Fetal Neonatal Med.* 2019;32(11):1835–1841.

.