Scientific Reports
https://doi.org/10.1038/s41598-026-44411-w

Article in Press

Prediction of hypertensive disorders of pregnancy in advanced-age pregnant women using SHAP value and XGBoost

Received: 2 April 2025
Accepted: 11 March 2026
Published online: 18 March 2026

Cite this article as: Wang J., Zhu H. & Gu W. Prediction of hypertensive disorders of pregnancy in advanced-age pregnant women using SHAP value and XGBoost. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-44411-w

Jue Wang, Hao Zhu & Weirong Gu

Prediction of hypertensive disorders of pregnancy in advanced-age pregnant women using SHAP value and XGBoost

Jue Wang[1†], Hao Zhu[1†], Weirong Gu[1*]

1. Department of Obstetrics, Obstetrics & Gynecology Hospital of Fudan University, Shanghai Key Lab of Reproduction and Development, Shanghai Key Lab of Female Reproductive Endocrine Related Diseases, 200433, Shanghai, China.

†These authors contributed equally to this study.

*Corresponding author: Weirong Gu, Department of Obstetrics, Obstetrics and Gynecology Hospital of Fudan University, Address: 419 Fangxie Road, Tel.: +86 21 33189900. Email: guweirong@fudan.edu.cn

Prediction of hypertensive disorders of pregnancy in advanced-age pregnant women using SHAP value and XGBoost

Abstract

Aim: To develop a cost-effective, predictive model for hypertensive disorders of pregnancy (HDP) in advanced-aged pregnant women based on demographic and lifestyle factors.

Methods: A large prospective, population-based, multicenter cohort study was conducted among advanced maternal-age pregnancies in China. Demographic and blood pressure data were collected from questionnaires of the first prenatal visits. The least absolute shrinkage and selection operator (Lasso) regression was applied for feature selection of risk factors, followed by XGBoost model construction and SHAP (SHapley Additive exPlanations) visualization.

Results: Lasso regression identified 9 risk factors, including systolic blood pressure in the first trimester (SBP1), diastolic blood pressure in the first trimester (DBP1), body mass index (BMI), family history of hypertension, multiparous parity, age, alcohol assumption, assisted reproductive technology (ART), and screen use. The XGBoost model was set with an optimized tune grid. The AUC of the model was 0.82, AUPRC of 0.41, with an accuracy of 0.88, sensitivity of 0.46, and specificity of 0.92. The SHAP demonstrated a novel predictive performance and clinical applicability.

Conclusion: The XGBoost-derived model offers a practical and simplified tool for individualized risk assessment in advanced maternal age pregnancies, facilitating early intervention and enhanced prenatal care.

Keywords: advanced maternal age; hypertension disorders of pregnancy; Lasso regression; XGBoost; SHAP;

Introduction

Hypertensive disorders of pregnancy (HDP), which include gestational hypertension, chronic hypertension, pre-eclampsia, and eclampsia, pose a significant global public health challenge and are a leading cause of maternal morbidity [1], [2] and mortality [3] across both high-income and low-income countries. These conditions are associated with adverse maternal and neonatal outcomes, including preterm labor, intrauterine growth restriction, and placental abruption. Early screening and prediction of HDP are critical for risk mitigation and enable timely intervention strategies to prevent severe maternal and perinatal complications.

Current evidence suggests that early screening for HDP [4] is beneficial. Existing approaches integrate clinical and demographic data, biochemical biomarkers, and sonographic techniques. Diverse combinations contribute to varying degrees of predictive accuracy. The integration of these methods enables a multilayered, tiered screening strategy that can be adapted to different clinical settings. The Fetal Medicine Foundation's combined screening algorithm is recommended by the International Society for the Study of Hypertension in Pregnancy [5] and the International Federation of Gynecology and Obstetrics [6] whenever feasible. However, these resource-intensive methods may not be practical for population-based screening [7]. Cost-effectiveness concerns and implementation challenges, particularly in regions with limited resources, often hinder its widespread adoption. Reliance on laboratory testing and ultrasound-based assessments may also restrict access to timely diagnosis and management in such areas.

Although current screening methods are effective, they frequently overlook dynamic, modifiable lifestyle factors that significantly influence maternal blood pressure. There is emerging evidence that highlights the profound impact of daily habits and environmental exposures on hypertension risk during pregnancy. This evidence underlines the need to incorporate these factors into risk assessment and prevention strategies. Key determinants include screen exposure [8], workload [9], [10], sleep duration [11], and physical activity [12], [13], which are increasingly recognized as regulators of blood pressure regulation during pregnancy. For instance, engaging in regular physical activity and managing one's weight before and during pregnancy can mitigate the risk of HDP [13], whereas high job demands and inadequate rest periods can result in insomnia, anxiety, and elevated blood pressure [14]. Furthermore, maternal self-efficacy in managing these lifestyle factors is essential for sustained behavioral change. Providing pregnant women with the knowledge and tools to adopt healthier habits can empower them to take control of their health and reduce their risk of HDP.

Advanced maternal age (typically defined as being 35 or older when giving birth) is increasingly recognized as a significant risk factor for adverse pregnancy outcomes, particularly HDP. In China, societal trends towards delayed childbearing and the promotion of third-child policies have contributed to a nearly 10% increase in advanced-age pregnancies over the past decade [15], making this a critical public health concern. Epidemiological studies demonstrate that pregnant women of advanced maternal age are at significantly higher risk of developing major obstetric complications, including preeclampsia, placental abruption, and postpartum hemorrhage. They are also more susceptible to other hypertensive disorders and related comorbidities [16], [17], [18]. Updated obstetric screening protocols are necessary to enable targeted prenatal surveillance and early intervention strategies in this vulnerable population.

Despite the significant influence of lifestyle factors on blood pressure during pregnancy, these elements remain underutilized in current HDP screening and prevention strategies. Unlike chronic hypertensive patients, who often initiate blood pressure monitoring, weight management, and pharmacological interventions preconceptionally or during early gestation, individuals with gestational hypertension and preeclampsia frequently exhibit lower adherence to lifestyle modifications. This discrepancy can be attributed to the absence of pregestational hypertension, resulting in an insufficient emphasis on lifestyle management and a tendency to underestimate the impact on disease progression. In order to address this gap in the literature, the present study aims to develop a novel, fully self-monitored predictive model for assessing the HDP risk in advanced maternal age pregnancy, based on easily accessible personal information. This innovative approach empowers pregnant women to actively, while providing a cost-effective, non-invasive tool for early risk identification.

Methods

Study population and design

This multi-centered prospective cohort study included women aged ≥ 35 years old pregnancies who gave birth in seven university hospitals across six provinces in China between 2015 and 2019. Pregnancies complicated by abortion, stillbirth, chronic hypertension, multiple pregnancies, aneuploidy, fetal structural malformations, or those lacking complete clinical data from the first and second trimester were excluded. All women were recruited at the first time of prenatal visit in the first trimester. Participants were recruited during their first prenatal visit in the first trimester, and all provided written informed consent before study enrollment. This study was conducted following the Declaration of Helsinki and received ethical approvals from all participating institutions (the leading institute of the project, approval No. 2016145-03).

Procedures

Upon enrollment, eligible patients agreed to complete a structured questionnaire at five predefined time points: early pregnancy (< 14 weeks), mid-term pregnancy (14-28 weeks), late pregnancy (28-36 weeks), at delivery, and during the postpartum period (6-12 weeks postnatal). This design allowed for longitudinal data collection throughout the study period. The questionnaire was administered by trained interviewers and included sociodemographic variables provided by participants, as well as clinical data (eg. blood pressure, maternal and neonatal outcomes) and laboratory results, which were recorded by clinical researchers. All data were systematically entered into a specialized computerized data management system. To ensure participant anonymity, each individual was assigned a unique identification number, which blinded personal information during data analysis. To enhance participant engagement and retention, each center appointed a dedicated quality controller responsible for verifying the logical consistency of recorded data and ensuring continuous follow-up. This structured approach facilitated rigorous data collection while maintaining participant confidentiality and study integrity.

Definitions of HDP

In this study, gestational hypertension and preeclampsia were combined as HDP due to limited statistical power for analyzing preeclampsia separately. Gestational hypertension was defined as systolic blood pressure (SBP) ≥ 140 mmHg and/or diastolic blood pressure (DBP) ≥ 90 mmHg after 20 weeks of gestation without proteinuria, in accordance with the clinical practice guideline of Chinese Society of Obstetrics and Gynecology 2020 [19]. Pre-eclampsia was characterized by hypertension developing after 20 weeks of gestation complicated with proteinuria (defined as 24 h urine protein ≥ 0.3 g or urine protein/creatinine ratio ≥ 0.3, or random urine protein ≥ 1+) or accompanied by evidence of end-organ damage in the absence of proteinuria. Chronic hypertension or pre-existing hypertension was excluded from the analysis, as it does not share a similar etiology with gestational hypertension or preeclampsia.

Variables included for analysis

All variables were systematically collected to characterize baseline demographics and maternal lifestyle behaviors for subsequent analysis. The following data were obtained from the self-reported questionnaires: (1) Baseline maternal characteristics were included age, pre-pregnancy BMI (calculated as pregestational body weight (kg) divided by the square of the body height (meter), multiparity (yes and no), Han ethnicity (yes or no), education level [basic (grade 1-5), secondary (grade 6-10), high school (grade 11-13), undergraduate, or graduate degree and above), annual household income (<30,000 yuan, 30,000 to 80,000 yuan, 80,000 to 120,000 yuan, 120,000 to 200,000 yuan, 200,000 to 300,000 yuan, or 300,000 yuan and more), natural conception (yes or no), pre-pregnancy diabetes mellitus (yes and no), family history of diabetes mellitus (yes and no), family history of hypertension (yes and no), pre-pregnancy thyroid abnormalities (yes and no), pre-pregnancy cardiac diseases (yes and no), pre-pregnancy renal diseases (yes and no), pre-pregnancy hepatic diseases (yes and no), pre-pregnancy immune-system diseases (yes and no), and pre-pregnancy hyperthyroidism (yes and no). (2) Lifestyle factor variables included smoking status[nonsmokers (never smoked before) or current smokers (smoked at least once a month from 6 months before pregnancy until enrollment)), alcohol habits[non-consumers or consumers (drank at least once a month in the previous six months before pregnancy until enrollment)), screen exposure[self-reported daily screen time (smartphones, computers, televisions, and pads) in the past week, classified as < 2 hours/day, 2-4 hours/day, 5-7 hours/day, or > 8 h hours/day), working time[mean hours spent on paid labor in the past week, classified as unemployment, < 10 hours/week, 10-40 hours/week, or > 40hours/week), physical exercise[active exercise in the past one week, classified as <0.5 hours/ day, 0.5-1 hours/day, 1-2 hours/day, or >2 hours/day) and sleep duration[total nightly sleep time in the past week, recorded as an integer value (hours)). (3) Blood pressure was measured at enrollment (<14 gestational weeks), second trimester (24-28 gestational weeks), third trimester (32-34 gestational weeks) and delivery. SBP1 denoted the systolic blood pressure in first trimester, and SBP2 denoted the second trimester, and so forth.

Statistical analysis

All candidate variables were subjected to a univariate analysis. Continuous variables were reported as means and standard deviations and compared using an independent Student's *t*-test if the data were normally distributed. Categorical variables were reported as numbers and percentages and compared using the chi-square test. Principal component analysis (PCA) was used to assess differences in participant constitution among the seven sites. One-quarter of the cases in each site were randomly selected. A heatmap was plotted to analyze the relationship between variables and blood pressure in each semester. To address missing data, stratified imputation was performed. Continuous and categorical variables were imputed using group-specific means and modes, respectively, stratified by hypertensive status. The cohort was randomly divided into a training set and a test set in an 8:2 ratio. The training set was used to establish the model, and the validation set was used to test the established model.

To identify the most predictive risk factors for HDPs, we utilized Lasso (Least Absolute Shrinkage and Selection Operator) regression for the final variable selection to mitigate information loss. Binary categorical variables were incorporated with the first level treated as the reference category. For ordered factors, the baseline reference level was set according to the result of the chi-square test. All categorical variables were converted into dummy variables before entering them into the Lasso regression model. Continuous variables, neither dummy variables, were centered to have a mean of zero and were scaled to unit variance to ensure a standardized feature matrix suitable for penalized regression. Collinearity was assessed using variance inflation factors (VIF), and a VIF of less than 5 indicated no collinearity among the variables.

To address the class imbalance in our dataset, we implemented a class weighting strategy during the XGBoost model training. This value was empirically determined to optimize the balance between sensitivity and specificity, with the specific aim of enhancing the model's ability to identify the minority class while maintaining reasonable overall performance. The model was optimized using the Area Under the Precision-Recall Curve (AUPRC) as the primary evaluation metric, which is particularly suitable for imbalanced classification problems. Sensitivity, specificity, and accuracy were determined from the optimal threshold using a confusion matrix. Internal validation was performed using 5-fold cross-validation to assess the model's generalizability and mitigate overfitting. To visualize the outcomes of our XGBoost model, we employed SHAP (SHapley Additive exPlanations) values to quantify the contribution of each feature to individual predictions. To further enhance the validation of our model, simple imputation was applied using the mean value for continuous variables and the mode for categorical factors. All statistical analyses were performed using SPSS (version 23.0, IBM) and RStudio (version 4.3.3) with a two-tailed test, and statistical significance was considered as $P < 0.05$.

Results

Demographics and clinical characteristics

A total of 13375 advanced-age pregnant women with a singleton were recruited in seven medical centers. After excluding pregnant women with chronic hypertension, structural malformation, aneuploidy, and loss to follow-up, 11339 eligible women met the inclusion criteria (Fig. 1). Table 1 described the demographic distribution and general characteristics of the participants included. The incidence of HDP in our study was 9.21%. The mean age and gestational age at the time of enrollment were 37.66 years and 12.62 weeks, and the mean BMI was 22.19 kg/m^2. Compared with non-HDP mothers, HDP mothers were more likely to be older and multiparous, to have higher BMI, lower household income, lower education level, and higher unemployment rate. University degree and above education level constituted 84.4% of the population recruited, and 24.9% of the families had an annual household income lower than 120,000 yuan. More women with HDP reported assisted reproductive technology. The baseline blood pressure at first entry was higher in HDP participants in both SBP and DBP. A family history of hypertension was present in 41.5% of the HDP compared to 30.1% of the non-HDP controls (P < 0.001).

Lifestyle factors associated with HDP

Participants without HDP smoked or consumed alcohol more frequently than HDP from 6 months preconception to enrollment. A total of 1.4 % of the women were current smokers, without significant differences between the cases and controls (P = 0.432, Table 1). More HDP patients reported unemployment with a gradual growth from 33.2% to 44.2% during pregnancy (Table 2) and the unemployment rate increased with the pregnancy stage. Participants with less than 4 hours of screen time every day show the highest prevalence of hypertension. HDP individuals reported slightly fewer amounts of screen time during the whole pregnancy than non-HDP, with 46.9% vs. 42.5% of participants using the screen for less than 5 hours respectively. However, exercise time was insignificant among the three trimesters.

A prediction model built based on lasso and XGBoost

Our PCA plot showed clustered differences between site 1-4 and site 5-7 (supplementary Fig. 1). The heatmap visualization revealed significant positive correlations (Fig. 2) between maternal BMI and blood pressure parameters across gestational periods: SBP and DBP in the first trimester (SBP1: Pearson's r = 0.204, P < 0.001, DBP1: r = 0.133, P < 0.001, supplementary Table 1) and second trimester (SBP2: r = 0.214, P < 0.001, DBP2: r = 0.142, P < 0.001). A weak negative correlation between workload1 and SBP1 (r = -0.046, P = 0.032) and DBP1 (r = -0.100, P < 0.001) in the first trimester. In the second trimester, both workload (r = -0.093, P < 0.001) and screen exposure time (screen2) (r = -0.086, P < 0.001) were negatively related to SBP2, but exercise time (exercise2) was positively related to DBP2 (r = 0.052, P < 0.013).

To rule out interference of blood pressure in the second trimester, we included baseline demographic, lifestyle factor and blood pressure in the first trimmest (Fig. 3). The lasso regression analysis selected 9 features, including systolic blood pressure in the first trimester (SBP1), diastolic blood pressure in the first trimester (DBP1), body mass index (BMI), family history of hypertension, multiparous parity, age, alcohol assumption, assisted reproductive technology (ART), and screen use.

The XGBoost model was set with an optimized tune grid. The model had an AUC of 0.82, an AUPRC of 0.41, an accuracy of 0.88, a sensitivity of 0.46, and a specificity of 0.92 (Fig. 4). The F1-score was 0.4. The 5-fold cross-validation demonstrated consistent model performance across all folds. The mean AUC-ROC was 0.74, and the AUPRC, which is more relevant for our imbalanced classification task, achieved a mean of 0.327.

SHAP as a tool for visualization of the XGBoost model

As a visual representation of the model, the SHAP was established to illustrate the impact of all influential factors on the incidence of HDP (Fig. 5). All factors were generally positively correlated with the risk of HDP. The most influential features in predicting hypertension were blood pressure in the first trimester and BMI, as indicated by their mean absolute SHAP values. Specifically, higher values of SBP in the first trimester and BMI consistently increase the predicted probability of hypertension. Two lifestyle factors, alcohol consumption and screen use, also had a positive impact, albeit a weak one.

Discussion

The present study successfully developed a fully predictive model for HDP by incorporating dynamic lifestyle factors, offering a practical approach for clinical application. Our findings indicated that prediction models based on demographic information and lifestyle factors can achieve clinically meaningful performance without relying on traditional biomarkers or hospital-based examinations. This represented a novel and cost-effective alternative to current HDP screening methods.

Traditional screening paradigms have focused on clinical biomarkers such as sFlt-1/PlGF [20] ratio (ratio of soluble fms-like tyrosine kinase-1 to placental growth factor) or uterine artery Doppler, which require laboratory results and specialized equipment. Our model, achieved an AUC of 0.80 using only clinical information and lifestyle parameters which clinically relevant predictive performance can be achieved without reliance on traditional biomarkers or specialized diagnostic tools. This approach opens new possibilities for population-level screening, particularly in resource-limited settings where healthcare access is restricted.

Our results challenge the assumption that reduced working hours are universally beneficial for maternal health. Specifically, individuals who work moderate hours (10-40 hours/week) or overtime (> 40 hours/week) exhibited a lower prevalence of hypertension compared to those with unemployed or worked fewer hours. This finding adds complexity to the understanding of work activity during pregnancy and its association with HDP. Limited research has directly examined the relationship between working hours and HDP, and existing studies report conflicting conclusions. Spinillo et al. [21] investigated the impact of work activity during pregnancy on the risk of severe preeclampsia. Women employed in clerical jobs had a significantly lower risk of developing severe preeclampsia compared to those who were unemployed at the start of pregnancy. Agarwal et al. [22] concluded working overtime was associated with higher odds of preeclampsia but not with gestational hypertension. Other studies have focused on the combined effects of shift work [23], [24] or prolonged standing or sedentary postures, rather than isolated working hours. Most of these studies noted that working-related physical or psychological demands are more consistently associated with adverse outcomes such as preterm birth and low birth weight infants [23], [24], [25]. Our findings challenge the simplistic assumption that "less work is better" and underscore the need for a nuanced analysis of how work activity influences maternal health outcomes.

Wage income represents another critical dimension of socioeconomic advantage. Unemployment is usually associated with reduced income and low education levels. Research examining individual-level socioeconomic status has demonstrated that women with higher socioeconomic status are less likely to develop pregnancy complications such as gestational hypertension and preeclampsia [26]. Specifically, adjusted odds ratios in these studies suggest that higher socioeconomic status may serve as a protective factor against these conditions. Our findings align with these conclusions. Furthermore, women receiving unemployment benefits exhibit higher rates of hypertensive disorders compared to those in work [27]. These results underscore the complex interplay between employment status, socioeconomic factors, and hypertensive disorders during pregnancy. In China, a substantial proportion choose to resign or take sick leave following pregnancy confirmation. This pattern highlights the need for a subtle interpretation of workplace policies. Rather than advocating for blanket recommendations such as reducing working hours, we propose tailored workplace accommodations based on individual job demands and maternal response. Targeted interventions aimed at improving socioeconomic conditions may effectively reduce the incidence of hypertensive disorders among pregnant women.

The predictive model demonstrated robust discriminative capacity, with an AUC of 0.81, indicating good overall performance. A particular strength was its high specificity of 0.9176, which translates to a low false-positive rate, making it an excellent tool for efficiently ruling out low-risk individuals in large-scale, community-level screening programs. However, the model's utility for identifying cases is limited by its modest sensitivity (0.46) and positive predictive value (PPV: 0.35). This suggests that the model is not appropriate for definitive diagnostic purpose, it serves a valuable role in self-screening and risk self-awareness. Women with elevated risk scores can be prompted to initiate closer self-monitoring (e.g., tracking blood pressure at home or ambulatory, modifying lifestyle factors) or seek professional evaluation, enabling earlier intervention if needed. Future iterations could refine specificity without compromising sensitivity by incorporating dynamic risk updates or integrating wearable device data for more precise behavioral assessment.

The strengths of the study lie in several key aspects. As a prospective multicenter study, it leverages a large-scale dataset covering mainland China, ensuring the diversity and representativeness of the sample, which provides a solid foundation for the generalizability of the findings. Second, the study employs an innovative methodology by integrating multidimensional data, such as lifestyle factors and clinical indicators, to construct a predictive model for HDP, offering a novel perspective for risk assessment. Additionally, this study is the first to focus on the impact of lifestyle factors on HDP and has revealed that moderate work during pregnancy may be beneficial for maternal health, providing new insights into preventive strategies for HDP. However, limitations also exist. The definition of lifestyle factors varies considerably across studies, and the heterogeneity in classification may compromise the external validity of the results. For instance, sedentary behavior is often conflated with activities such as watching television or scrolling the internet, while physical exercise is sometimes combined with occupation-related aerobic workload. This lack of standardized definitions poses challenges in drawing consistent conclusions regarding lifestyle factors' impact on pregnancy health outcomes. On the other hand, while our model successfully incorporated lifestyle variables as non-invasive predictors of HDP, we must emphasize that their contributions to risk stratification were relatively modest compared to other variables, such as blood pressure at the first visit.

Conclusion

In summary, our study presented the relationship between work-life balance and maternal health outcomes in pregnant women. Daily routines and behaviors that individuals consistently engage in may influence maternal blood pressure. Contrary to previous hypotheses, our findings suggest that overtime work during pregnancy does not appear to pose a significant risk to hypertensive women. However, further study is needed to clarify the mechanisms underlying the protective effect of these behaviors.

Statements and Declarations

Funding

This study was supported by the Shanghai Medical Research Program of the Science and Technology Innovation Action Plan (No. 21Y11907800) and the Shanghai Municipal Health Commission (No. 202240082).

Competing interests

The authors declare they have no financial interests.

Author contribution

W. Gu contributed to the study's design and provided the original data. J. Wang analyzed the data and wrote the first draft of the manuscript. H. Zhu revised the manuscript. All authors have read and approved the final manuscript.

Ethics approval for the study

The study was approved by the Peking University Third Hospital (the leading hospital of the project) Medical Science Research Ethics Committee (No. 145-03, 2016). All participants were recruited from the following hospitals: the Peking University Third Hospital, Peking University First Hospital, Obstetrics and Gynecology Hospital of Fudan University, West China Second University Hospital of Sichuan University, the Third Affiliated Hospital of Guangzhou Medical University, Tongji Hospital affiliated to Tongji Medical College of Huazhong University of Science & Technology, the First Affiliated Hospital of Chongqing Medical University. Each participating center obtained independent ethical approval from their respective institutional review boards before patient enrollment.

Consent to participate and publish

Written informed consent was obtained from all participants for the publication of their anonymized data and any accompanying images or case details.

Availability of data and material

The datasets generated and analyzed during the current study are not publicly available due to ongoing research using the same cohort but are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used DeepSeek to assist with language refinement and grammatical error correction. The tool enhanced clarity, improved phrasing, and ensured linguistic accuracy throughout the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References:

[1] L. Say *et al.*, "Global causes of maternal death: a WHO systematic analysis," *Lancet Glob. Health*, vol. 2, no. 6, pp. e323–e333, Jun. 2014, doi: 10.1016/S2214-109X(14)70227-X.

[2] L. Jiang *et al.*, "A global view of hypertensive disorders and diabetes mellitus during pregnancy," *Nat. Rev. Endocrinol.*, vol. 18, no. 12, pp. 760–775, Dec. 2022, doi: 10.1038/s41574-022-00734-y.

[3] N. J. Kassebaum *et al.*, "Global, regional, and national levels and causes of maternal mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 384, no. 9947, pp. 980–1004, Sep. 2014, doi: 10.1016/S0140-6736(14)60696-6.

[4] P. Chaemsaithong, D. S. Sahota, and L. C. Poon, "First trimester preeclampsia screening and prediction," *Am. J. Obstet. Gynecol.*, vol. 226, no. 2, pp. S1071-S1097.e2, Feb. 2022, doi: 10.1016/j.ajog.2020.07.020.

[5] L. A. Magee *et al.*, "The 2021 International Society for the Study of Hypertension in Pregnancy classification, diagnosis & management recommendations for international practice," *Pregnancy Hypertens.*, vol. 27, pp. 148–169, Mar. 2022, doi: 10.1016/j.preghy.2021.09.008.

[6] L. C. Poon *et al.*, "The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention," *Int. J. Gynecol. Obstet.*, vol. 145, no. S1, pp. 1–33, May 2019, doi: 10.1002/ijgo.12802.

[7] P. Wu, M. Green, and J. E. Myers, "Hypertensive disorders of pregnancy," *BMJ*, p.

e071653, Jun. 2023, doi: 10.1136/bmj-2022-071653.

[8] S. Tang, Y. Liu, Y. Gu, Q. Yang, and Q. Wang, "Effects of electronic screen exposure time on hypertensive disorders in pregnancy: a retrospective cohort study," *BMC Public Health*, vol. 24, no. 1, p. 1367, May 2024, doi: 10.1186/s12889-024-18793-3.

[9] T. Tatarian *et al.*, "Impact of Surgeon Workload on Pregnancy Outcomes in Surgical Trainees and Faculty," *Ann. Surg.*, Jun. 2024, doi: 10.1097/SLA.0000000000006426.

[10] E. Mozurkewich, "Working conditions and adverse pregnancy outcome: a meta-analysis," *Obstet. Gynecol.*, vol. 95, no. 4, pp. 623–635, Apr. 2000, doi: 10.1016/S0029-7844(99)00598-0.

[11] Q. Lu *et al.*, "Sleep disturbances during pregnancy and adverse maternal and fetal outcomes: A systematic review and meta-analysis," *Sleep Med. Rev.*, vol. 58, p. 101436, Aug. 2021, doi: 10.1016/j.smrv.2021.101436.

[12] K. Melzer, Y. Schutz, M. Boulvain, and B. Kayser, "Physical Activity and Pregnancy: Cardiovascular Adaptations, Recommendations and Pregnancy Outcomes," *Sports Med.*, vol. 40, no. 6, pp. 493–507, Jun. 2010, doi: 10.2165/11532290-000000000-00000.

[13] M. F. Mottola *et al.*, "No. 367-2019 Canadian Guideline for Physical Activity throughout Pregnancy," *J. Obstet. Gynaecol. Can.*, vol. 40, no. 11, pp. 1528–1537, Nov. 2018, doi: 10.1016/j.jogc.2018.07.001.

[14] H. E. Miller *et al.*, "Associations between anxiety, sleep, and blood pressure

parameters in pregnancy: a prospective pilot cohort study," *BMC Pregnancy Childbirth*, vol. 24, no. 1, p. 366, May 2024, doi: 10.1186/s12884-024-06540-w.

[15] J. Cao *et al.*, "Trends in maternal age and the relationship between advanced age and adverse pregnancy outcomes: a population-based register study in Wuhan, China, 2010–2017," *Public Health*, vol. 206, pp. 8–14, May 2022, doi: 10.1016/j.puhe.2022.02.015.

[16] M.-L. Tian *et al.*, "The Effect of 2016 Chinese second-child policy and different maternal age on pregnancy outcomes in Hebei Province, China," *BMC Pregnancy Childbirth*, vol. 23, no. 1, p. 267, Apr. 2023, doi: 10.1186/s12884-023-05552-2.

[17] H. Zhu *et al.*, "Trajectories tracking of maternal and neonatal health in eastern China from 2010 to 2021: A multicentre cross-sectional study," *J. Glob. Health*, vol. 14, p. 04069, Mar. 2024, doi: 10.7189/jogh.14.04069.

[18] Y. Zhou *et al.*, "Association of maternal age with adverse pregnancy outcomes: A prospective multicenter cohort study in China," *J. Glob. Health*, vol. 13, p. 04161, Dec. 2023, doi: 10.7189/jogh.13.04161.

[19] Hypertensive Disorders in Pregnancy Subgroup, Chinese Society of Obstetrics and Gynecology, Chinese Medical Association, "[Diagnosis and treatment of hypertension and pre-eclampsia in pregnancy: a clinical practice guideline in China （2020）]," *Zhonghua Fu Chan Ke Za Zhi*, vol. 55, no. 4, pp. 227–238, Apr. 2020, doi: 10.3760/cma.j.cn112141-20200114-00039.

[20] H. Stepan *et al.*, "Clinical utility of sFlt-1 and PlGF in screening, prediction, diagnosis and monitoring of pre-eclampsia and fetal growth restriction," *Ultrasound*

Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol., vol. 61, no. 2, pp. 168–180, Feb. 2023, doi: 10.1002/uog.26032.

[21] A. Spinillo, E. Capuzzo, L. Colonna, G. Piazzi, S. Nicola, and F. Baltaro, "The Effect of Work Activity in Pregnancy on the Risk of Severe Preeclampsia," *Aust. N. Z. J. Obstet. Gynaecol.*, vol. 35, no. 4, pp. 380–385, Nov. 1995, doi: 10.1111/j.1479-828X.1995.tb02146.x.

[22] I. Agarwal *et al.*, "First-trimester occupational exposures and hypertensive disorders of pregnancy among US nurses," *Occup. Med.*, vol. 73, no. 7, pp. 419–426, Oct. 2023, doi: 10.1093/occmed/kqad089.

[23] C. Cai *et al.*, "The impact of occupational shift work and working hours during pregnancy on health outcomes: a systematic review and meta-analysis," *Am. J. Obstet. Gynecol.*, vol. 221, no. 6, pp. 563–576, Dec. 2019, doi: 10.1016/j.ajog.2019.06.051.

[24] N. Suzumori *et al.*, "Effects of long working hours and shift work during pregnancy on obstetric and perinatal outcomes: A large prospective cohort study—Japan Environment and Children's Study," *Birth*, vol. 47, no. 1, pp. 67–79, Mar. 2020, doi: 10.1111/birt.12463.

[25] K. T. Palmer, M. Bonzini, E. C. Harris, C. Linaker, and J. P. Bonde, "Work activities and risk of prematurity, low birth weight and pre-eclampsia: an updated review with meta-analysis," *Occup. Environ. Med.*, vol. 70, no. 4, pp. 213–222, Apr. 2013, doi: 10.1136/oemed-2012-101032.

[26] L. Nicholls-Dempsey, A. Badeghiesh, H. Baghlaf, and M. H. Dahan, "How does high

socioeconomic status affect maternal and neonatal pregnancy outcomes? A population-based study," *Am. J. Obstet. Gynecol.*, vol. 226, no. 1, pp. S452–S453, Jan. 2022, doi: 10.1016/j.ajog.2021.11.751.

[27] M. Morales-Suárez-Varela *et al.*, "Unemployment and pregnancy outcomes: A study within the Danish National Birth Cohort," *Scand. J. Public Health*, vol. 39, no. 5, pp. 449–456, Jul. 2011, doi: 10.1177/1403494811407672.

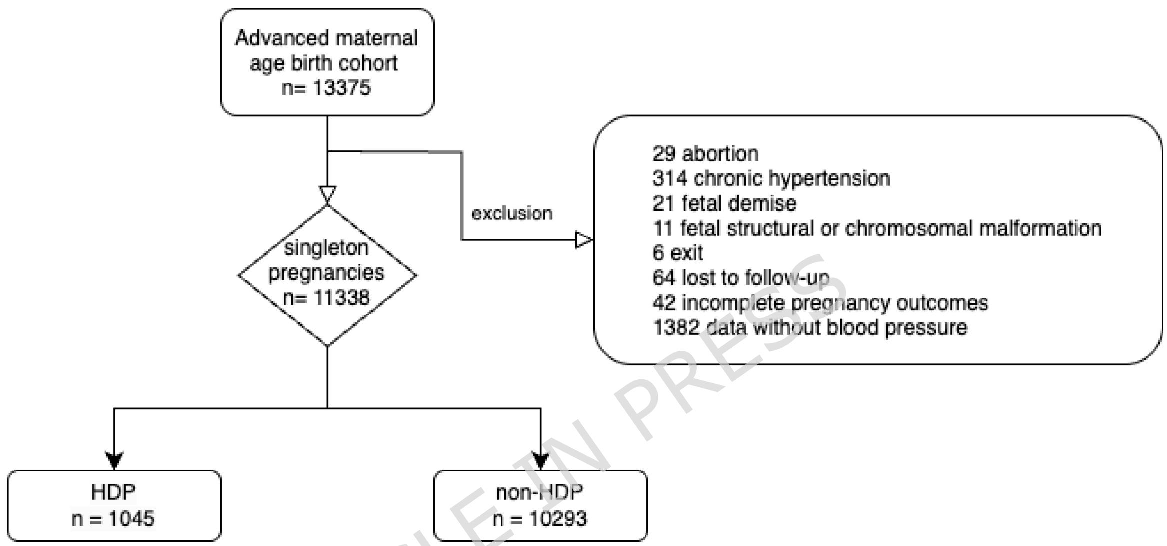
Fig. 1. The flow chart of participants included.

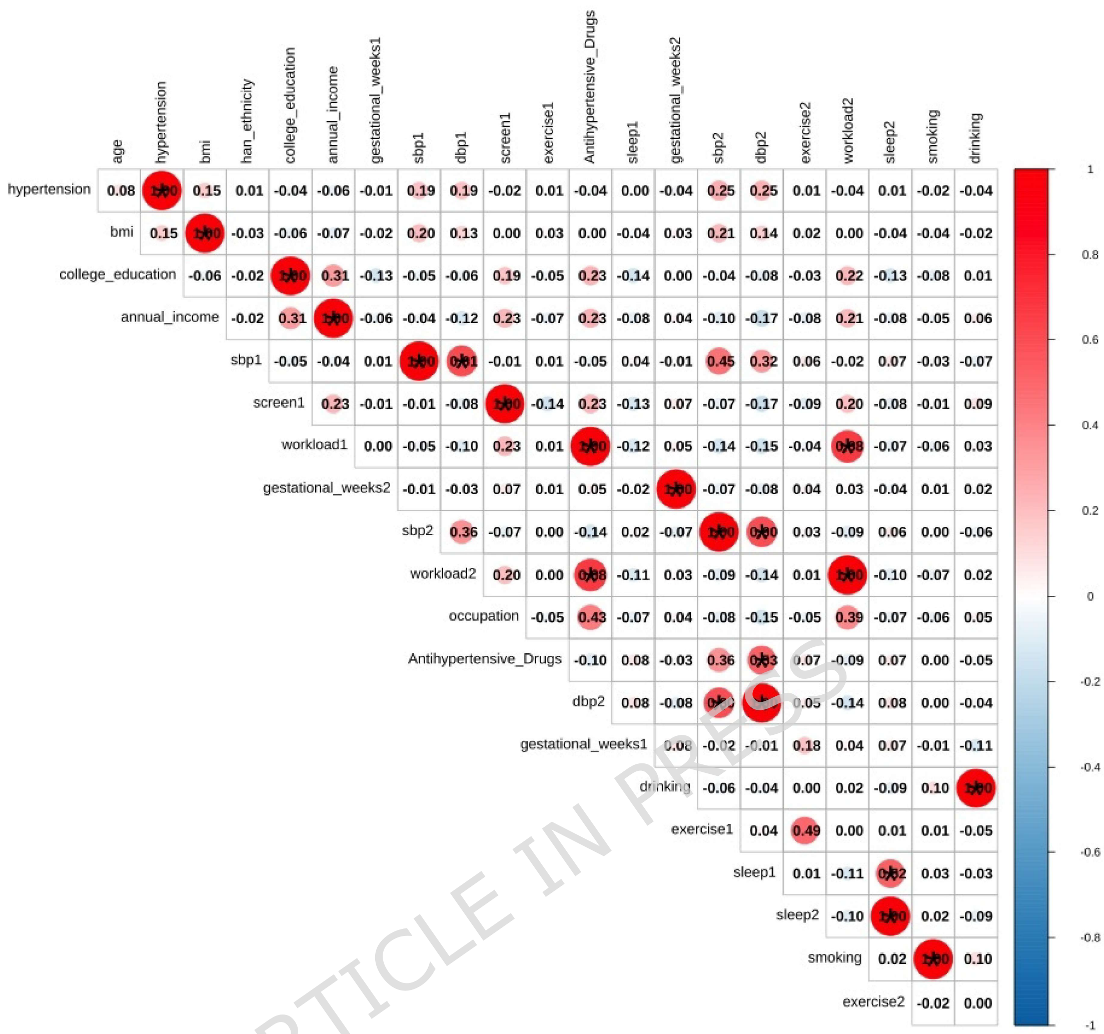
Fig. 2. The heatmap which visualized the correlation of risk factors and blood pressure in the first and second trimesters.

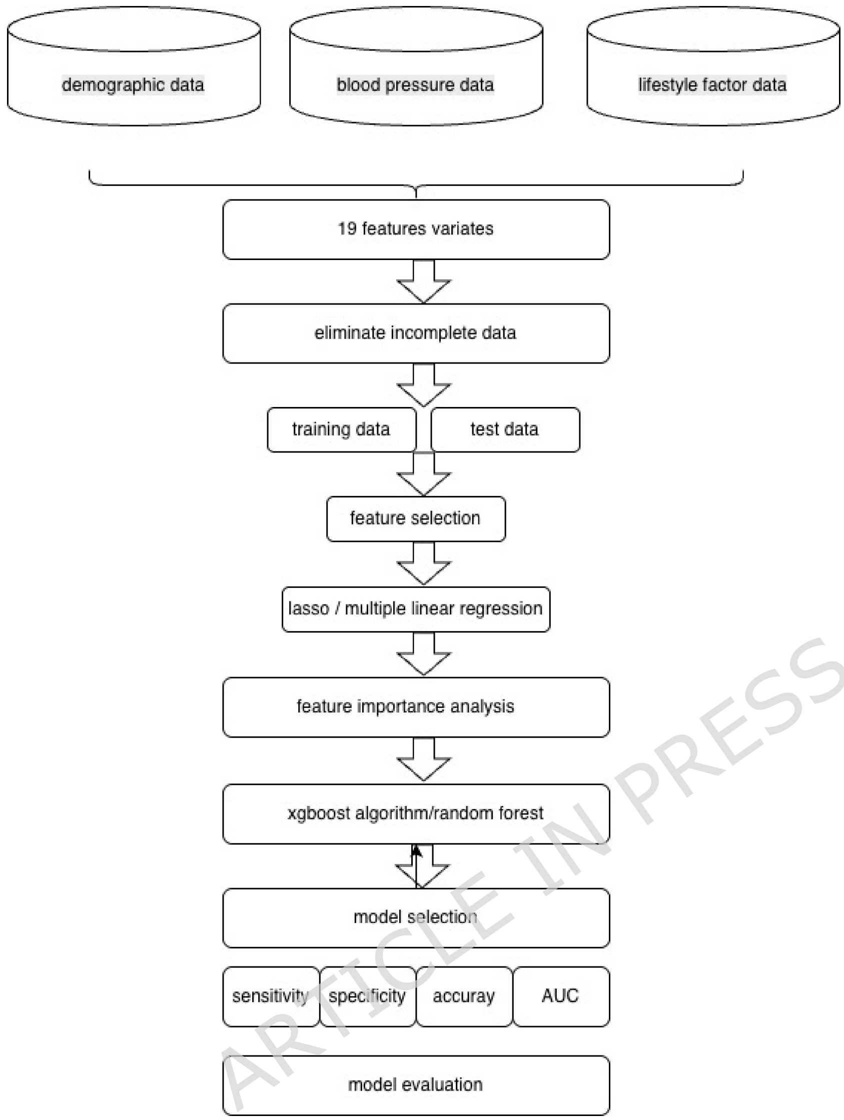
Fig. 3. The statistical analytic process of building the model.

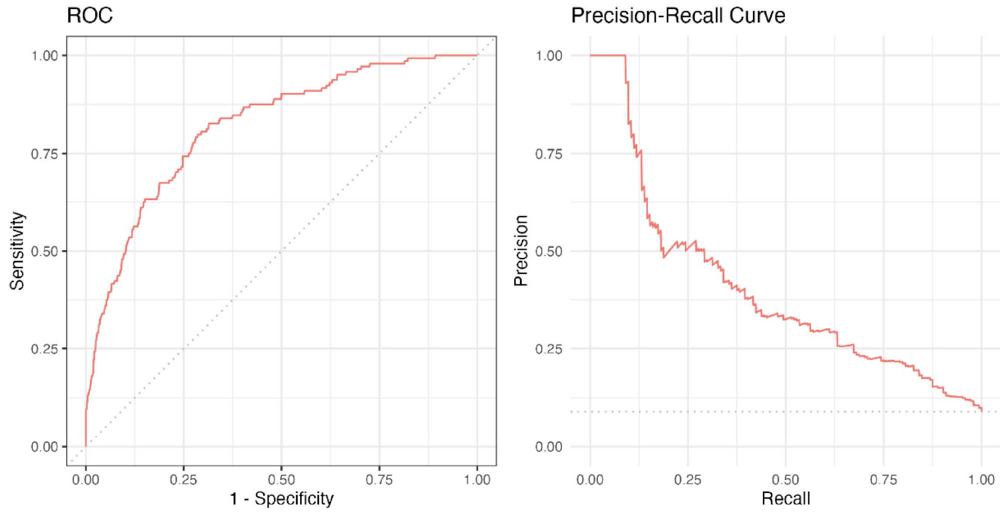
Fig. 4. AUC and AUPRC of the XGBoost model.

Fig. 5. SHAP for predicting the risk of HDP in advanced-age pregnant women.






demographic data
blood pressure data
lifestyle factor data
19 features variates
eliminate incomplete data
training data
test data
feature selection
lasso / multiple linear regression
feature importance analysis
xgboost algorithm/random forest
model selection
sensitivity
specificity
accuray
AUC
model evaluation


ROC
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
Precision-Recall Curve
Precision
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Recall

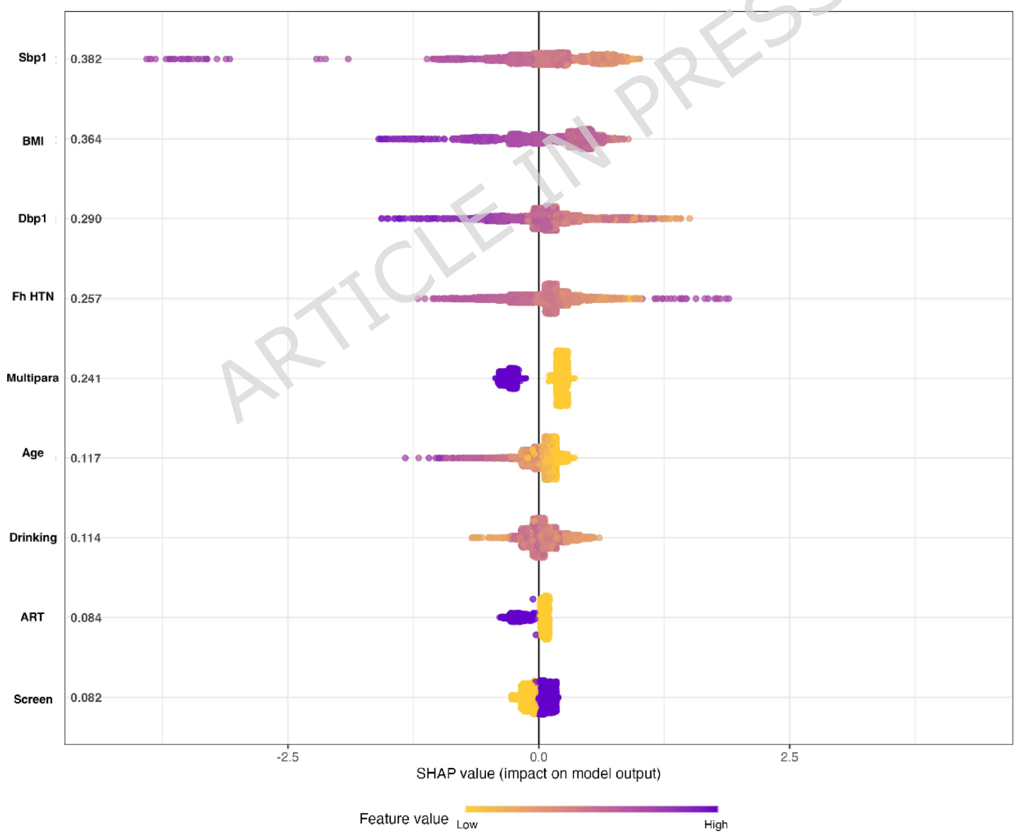

Sbp1 0.382
BMI 0.364
Dbp1 0.290
Fh HTN 0.257
Multipara 0.241
Age 0.117
Drinking 0.114
ART 0.084
Screen 0.082
-2.5
0.0
2.5
SHAP value (impact on model output)
Feature value
Low
High

Table 1. The demographic characteristics of included participants.

	non-HDP	HDP	*P* value
	N = 10293	N = 1045	
Maternal age, mean±SD, year	37.63±2.31	38.00±2.60	<0.001*
Gestation at entry, mean±SD, week	12.63±3.64	12.56±4.20	0.487*
BMI, mean±SD, kg/m^2	22.03±2.86	23.75±3.62	<0.001*
Primiparity, n(%)	3903(37.9)	490(46.9)	<0.001
Sbp, mean±SD, mmHg	113.45±10.67	120.65±9.83	<0.001
Dbp, mean±SD, mmHg	70.74±8.66	76.15±8.38	<0.001
Han nationality, n(%)	9523(96.9)	966(96.5)	0.501
Education, n(%)			<0.001
Basic (grade 1-5)	74(0.8)	15(1.5)	
Secondary (grade 6-10)	471(4.8)	65(6.6)	
High school (grade 11-13)	924(9.5)	122(12.3)	
University	6186(63.5)	626(63.1)	
Graduate and above	2091(21.4)	164(1`6.5)	
Unemployment, n(%)	1262(12.9)	159(15.9)	0.037
Smoking, n(%)	140(1.5)	11(1.1)	0.432
Alcohol use, n(%)	753(7.8)	54(5.5)	**0.010**
Passive smoking, n(%)	1330(14.1)	134(14.1)	0.053
Husband smoking, n(%)	2620(28.3)	285(30.5)	0.221
Husband's alcohol use, n(%)	3089(33.4)	303(32.5)	0.584
Annual income, RMB, n(%)			**0.003**
<30,000	185(2)	24(2.6)	
30,000-80,000	626(6.8)	79(8.5)	
80,000-120,000	1452(15.7)	169(18.1)	
120,000-200,000	2264(24.5)	250(26.8)	
200,000-300,000	2159(23.4)	191(20.4)	
≥300,000	2560(27.7)	221(23.7)	
Fertilization, n(%)			<0.001
Natural conception	8215(84)	773(77.6)	
IVF	1422(14.5)	195(19.6)	
IUI	26(0.3)	7(0.7)	
Ovarian stimulation,	72(0.7)	13(1.3)	
ICSI	17(0.2)	2(0.2)	
Others	31(0.3)	6(0.6)	
Family history of HTN, n(%)	2943(30.1)	411(41.5)	<0.001
Family history of diabetes, n(%)	1410(14.4)	170(17.2)	0.019
PGDM, n(%)	99(1.0)	27(2.7)	<0.001
Pre-pregnancy cardiac disease	119(1.2)	22(2.2)	**0.032**
Pre-pregnancy renal disease	105(1.1)	13(1.3)	0.796
Pre-pregnancy hepatic disease	410(4.2)	37(3.7)	0.603

Pre-pregnancy immune system disease	179(1.8)	29(2.9)	**0.047**
Pre-pregnancy hyperthyroidism	113(1.2)	11(1.1)	0.909

* Kruskal-Wallis; Sbp: systolic blood pressure; Dbp: diastolic blood pressure; IVF: in vitro fertilization; IUI: Intrauterine insemination; ICSI: Intracytoplasmic sperm injection; HTN: hypertension; PGDM: pre-gestational diabetes mellitus;

Table 2. Lifestyle comparison in early, mid-term and late pregnancy.

	Early trimester			Mid trimester			Late trimester		
	non-HDP	HDP	*P* value	non-HDP	HDP	*P* value	non-HDP	HDP	*P* value
Screen, h/day			**0.031**			**0.005**			**0.004**
<2, n(%)	972(10.5)	117(12.5)		762(8.3)	85(9.2)		698(7.7)	98(10.8)	
2-4, n(%)	2966(32)	321(34.4)		2655(28.8)	307(33.2)		2778(30.6)	285(31.5)	
5-7, n(%)	3262(35.3)	318(34)		3702(40.2)	359(38.8)		3826(42.2)	367(40.6)	
>8, n(%)	2055(22.2)	178(19.1)		2100(22.8)	174(18.8)		1766(19.5)	155(17.1)	
Workload, h/week			0.072			**0.029**			**0.007**
Unemployment	2673(28.9)	310(33.2)		2889(31.4)	327(35.4)		3480(38.4)	400(44.2)	
<10, n(%)	870(9.4)	85(9.1)		556(6)	63(6.8)		612(6.7)	67(7.4)	
10-40, n(%)	3318(35.8)	324(34.7)		3863(42)	378(40.9)		3814(42.1)	340(37.5)	
>40, n(%)	2399(25.9)	215(23)		1892(20.6)	157(17.0)		1163(12.8)	99(10.9)	
Physical exercise, h/day			0.236			0.332			0.292
<0.5, n(%)	3382(36.5)	320(34.3)		2458(26.7)	271(29.3)		2641(29.1)	283(31.2)	
0.5-1, n(%)	3927(42.4)	419(44.9)		4600(49.9)	450(48.6)		4489(49.5)	452(49.9)	
1-2, n(%)	1585(17.1)	151(16.2)		1707(18.5)	158(17.1)		1564(17.2)	138(15.2)	
>2, n(%)	361(3.9)	44(4.7)		453(4.9)	47(5.1)		375(4.1)	33(3.6)	
Sleep duration, h/day	7.94 ± 1.04	7.94 ± 1.10	0.800	7.79 ± 1.10	7.82 ± 0.95	0.352	7.64 ± 0.95	7.68 ± 1.06	0.299