Endocrine Journal 2025, 72 (4), 387-398. doi:10.1507/endocrj.EJ24-0449

ORIGINAL

# Prediction of insulin resistance using multiple adaptive regression spline in Chinese women

Shih-Peng Mao[1)], Chen-Yu Wang[2),3)], Chi-Hao Liu[4),5)], Chung-Bao Hsieh[6)], Dee Pei[7)], Ta-Wei Chu[2),8)] and Yao-Jen Liang[3)]

[1)] *Department of Obstetrics and Gynecology, Shuang Ho Hospital, Taipei Medical University, New Taipei City 235, Taiwan, R.O.C.*
[2)] *Department of Obstetrics and Gynecology, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan, R.O.C.*
[3)] *Graduate Institute of Applied Science and Engineering, Fu Jen Catholic University, New Taipei City 242, Taiwan, R.O.C.*
[4)] *Division of Nephrology, Department of Medicine, Kaohsiung Armed Forces General Hospital, Kaohsiung 802, Taiwan, R.O.C.*
[5)] *School of Medicine, National Defense Medical Center, Taipei 114, Taiwan, R.O.C.*
[6)] *Department of Surgery, Kaohsiung Armed Forces General Hospital, Kaohsiung 802, Taiwan, R.O.C.*
[7)] *Division of Endocrinology and Metabolism, Department of Internal Medicine, Fu Jen Catholic University Hospital, School of Medicine, College of Medicine, Fu Jen Catholic University, New Taipei City 242, Taiwan, R.O.C.*
[8)] *MJ Health Research Foundation, Taipei 114, Taiwan, R.O.C.*

**Abstract.** Insulin resistance (IR) is the core for type 2 diabetes and metabolic syndrome. The homeostasis assessment model is a straightforward and practical tool for quantifying insulin resistance (HOMA-IR). Multiple adaptive regression spline (MARS) is a machine learning method used in many research fields but has yet to be applied to estimating HOMA-IR. This study uses MARS to build an equation to estimate HOMA-IR in pre-menopausal Chinese women based on a sample of 4,071 healthy women aged 20–50 with no major diseases and no medication use for blood pressure, blood glucose or blood lipids. Thirty variables were applied to build the HOMA-IR model, including demographic, laboratory, and lifestyle factors. MARS results in smaller prediction errors than traditional multiple linear regression (MLR) methods, and is thus more accurate. The model was established based on key impact factors including waist-hip ratio (WHR), C reactive protein (CRP), uric acid (UA), total bilirubin (TBIL), leukocyte (WBC), serum glutamic oxaloacetic transaminase (GOT), high-density lipoprotein cholesterol (HDL-C), systolic blood pressure (SBP), serum glutamic pyruvic transaminase (GPT), and triglycerides (TG). The equation is as following:

HOMA-IR = 6.634 – 1.448MAX(0, 0.833 – WHR) + 10.152MAX(0, WHR – 0.833) – 1.351MAX(0, 0.7 – CRP) – 0.449MAX(0, CRP – 0.7) + 1.062MAX(0, UA – 8.5) + +1.047(MAX(0, 0.83 – TBIL) + 0.681MAX(0, WBC – 11.53) – 0.071MAX(0, 11.53 – WBC) + 0.043MAX(0, 24 – GOT) – 0.017MAX(0, GOT – 24) + 0.021MAX(0, 59 – HDL) – 0.005MAX(0, HDL – 59) – 0.013MAX(0, 141 – SBP) – 0.033MAX(0, 100 – GPT) + 0.013MAX(0, GPT – 100) – 0.004MAX(303 – TG)

Results indicate that MARS is a more precise tool than fasting plasma insulin (FPI) levels, and could be used in the daily practice, and further longitudinal studies are warranted.

*Key words*: Insulin resistance, Homeostasis assessment model, Multiple adaptive regression spline

Submitted Aug. 28, 2024; Accepted Dec. 3, 2024 as EJ24-0449
Released online in J-STAGE as advance publication Feb. 1, 2025
Correspondence to: Yao-Jen Liang, PhD, Graduate Institute of Applied Science and Engineering, Fu Jen Catholic University, No.510, Zhongzheng Rd., Xinzhuang Dist., New Taipei City 242, Taiwan, R.O.C.
E-mail: men72227@gmail.com

Appendix: IR, insulin resistance; MARS, multiple adaptive regression spline; MLR, multiple linear regression; FPI, fasting plasma insulin; FPG, fasting plasma glucose; T2D, type 2 diabetes; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; WHR, waist-hip ratio; WBC, leukocyte; Hb, hemoglobin; Plt, platelets; TBIL, total bilirubin; Alb, albumin; Glo, globulin; ALP, alkaline phosphatase; GOT, serum glutamic oxaloacetic transaminase; GPT, serum glutamic pyruvic transaminase; γ-GT, serum γ-glutamyl transpeptidase; LDH, lactate dehydrogenase; Cr, creatinine; UA, uric acid; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; Ca, plasma calcium concentration; P, plasma phosphate concentration; TSH, thyroid stimulating hormone; CRP, C reactive protein; MS, marital status; IL, income level; Edu, education level; SH, sleep hours; SMAPE, symmetric mean absolute percentage error; RAE, relative absolute error; RRSE, root relative squared error; RMSE, root mean squared error.

## Background

The prevalence of type 2 diabetes (T2D) has increased dramatically in recent years, with total global cases increasing from 462 million in 2017 to nearly 700 million in 2023, an increase of over 50% in just six years [1]. With nearly 1 million deaths per year, T2D is currently the world's 9th leading cause of death. Similar trend is found in Taiwan, where around 9.7% of the population had T2D in 2021 [2]. T2D is the 5th leading cause of death in Taiwan, where T2D and its comorbidities account for 11.5% of total expenses by the government's national health insurance scheme [3]. Thus, early diagnosis and appropriate management of T2D is a critical issue both in Taiwan and globally.

The underlying pathophysiologies of T2D are increased insulin resistance (IR) and decreased insulin secretion [4], and IR is simply defined as a "blunting of insulin's hypoglycemic effects" [5]. There are many different methods to measure IR. The most sophisticated method involves the use of a "euglycemic clamp," wherein the amount of glucose water infused to maintain plasma glucose level in the normal range for 2 hours would be the IR value, and a higher glucose infused volume is negatively correlated with IR. However, this approach is expensive and labor intensive to perform. Other tests such as insulin suppression test, intravenous glucose tolerance test and oral glucose tolerance test all have their unique characteristics and have been developed and applied in different studies. Of these tests, the hemostasis assessment model for insulin resistance (HOMA-IR) was first reported by Matthews *et al.* [6], calculating IR through multiple measurements of fasting plasma glucose (FPG) and insulin (FPI) with a coefficient. Because of its simplicity, HOMA-IR has been extensively used.

Many studies have attempted to identify and characterize various impact factors for IR, including inflammation, obesity, fatty liver, metabolic syndrome, and polycystic ovary syndrome. Aging is also considered to be related to IR [7], reporting that severity of IR increases with age.

However, the findings presented by Barbieri *et al.* suggest that IR and reduced insulin secretion are not necessarily correlated with age [8].

Most of the aforementioned studies used traditional statistical methods. Recently, however, the development of artificial intelligence and related machine learning techniques have been widely applied in medical research. Among these methods, multivariate adaptive regression spline (MARS) could provide a specific equation which captures and makes explicit non-linear relationships between variables, allowing MARS to outperform traditional multiple linear regression (MLR) approaches in predicting the behavior of dependent variables. The present study uses MARS in building an equation based on demographic, biochemistry, and lifestyle data for the prediction of HOMA-IR in healthy Chinese women.

## Methods

### *Participant and Study Design*

The present study followed our previously published protocol [9]. Data were sourced from the Taiwan MJ cohort, an ongoing prospective cohort of health examinations conducted by the MJ Health Screening Centers in Taiwan [10]. These examinations cover more than 100 important biological indicators, including anthropometric measurements, blood tests, imaging tests, *etc.* Each participant completed a self-administered questionnaire to collect information related to personal and family medical history, current health status, lifestyle, physical exercise, sleep habits, and dietary habits [11]. It should be noted that this study is a secondary data analysis. The data of our study participants were obtained from health checkups conducted at the MJ Clinic, with general consent obtained for future anonymous studies. All or part of the data used in this research were authorized by and received from MJ Health Research Foundation (Authorization Code: MJHRF2023007A). Any interpretations or conclusions described in this paper do not represent the views of MJ Health Research Foundation. Please refer to the technical report annually [11]. The study protocol was approved by the Institutional Review Board of the Kaohsiung Armed Forces General Hospital (IRB No.: KAFGHIRB 112-006). Since no samples were collected from patients, a short IRB review was approved and no consent was required. The data set initially included 646,728 healthy participants. Filtering for the exclusion criteria listed below, the final sample for analysis included 4,071 subjects (see Fig. 1).

The study exclusion criteria were:

1. Male;
2. Age <20 and >50 years old;
3. Taking medications for metabolic syndrome at the time of the study
4. Missing data of components of metabolic syndrome

On the day of the study, a senior MJ nursing staff member recorded the subject's medical history, including information on any current medications, and performed a physical examination. Waist circumference was measured horizontally at the level of the natural waist. Body mass index (BMI) was calculated as the subject's body weight (kg) divided by the square of the subject's height (m). Both systolic blood pressure and diastolic blood pressure were measured using a standard mercury

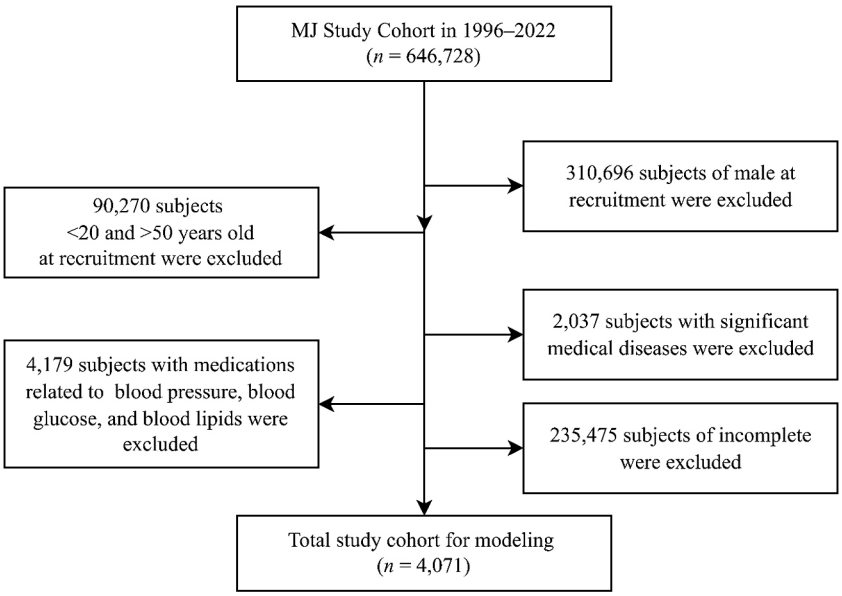



**Fig. 1** Flowchart of sample selection from the MJ study cohort

sphygmomanometer on the right arm while seated.

After fasting for 10 hours, blood samples were drawn for biochemical analysis. Plasma was separated from the blood within 1 hour of collection and stored at 30°C until analysis for FPG and lipid profiles. FPG was measured using a glucose oxidase method (YSI 203 glucose analyzer, Yellow Springs Instruments, Yellow Springs, USA). Total cholesterol and triglyceride (TG) levels were measured using a dry, multilayer analytical slide method with the Fuji Dri-Chem 3000 analyzer (Fuji Photo Film, Tokyo, Japan). Serum high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) concentrations were analyzed using an enzymatic cholesterol assay, following dextran sulfate precipitation. Beckman Coulter AU 5800 biochemical analyzer determined the urine microalbumin by turbidimetry.

The 30 variables of the present study are shown in Table 1 (dependent variables) which could be grouped into three group categories, demographics, biochemistry, and lifestyle and HOMA-IR, of the independent variable. It should be noted that fasting plasma glucose and insulin levels were not included in the statistical analysis since HOMA-IR was calculated from these two parameters.

### *Equation used to calculate HOMA-IR*

$$\text{HOMA}-\text{IR} = \frac{\text{FPI}(\mu\text{U/mL}) \times \text{FPG}(\text{mg/dL})}{405}$$

### *Traditional statistics*

To compare HOMA-IR performance using different demographic, biochemistry, and lifestyle parameters, we used *t*-tests to compare HOMA-IR for marital status, and analyzed variances to compare the HOMA-IR in terms of ordinal data such as education and income level (see Table 2). A simple correlation was applied to evaluate the relationship between HOMA-IR and other parameters (see Table 3). Considering the potential multicollinearity between WBC and CRP, both of which are inflammatory markers, principal component analysis (PCA) was applied to evaluate the impact of these two factors on HOMA-IR.

The aforementioned tests were performed using SPSS version 19.0 (IBM Inc., Armonk, New York).

### *Machine learning method*

The present study used the multiple adaptive regression spline (MARS) which is unique in that it provides an equation. For each independent variable (x), there might be a non-linear relationship. MARS can capture these relationships, and thus, is expected to provide more accurate results than the traditional multiple regression.

A PubMed search shows that MARS has been extensively used in other studies that can be classified as follows: 1. Comparison between different AI and MARS [12, 13]. 2. Demonstration of the method of MARS itself [14, 15]. 3. Animal studies [16]. 4. Genetic studies [17]. 5. Medical research [18-20]. The present study is one of the first to apply MARS to the study of IR. MARS is useful for developing adaptable models suited for high-dimensional data. This modeling method uses an expansion structure reliant on product spline basis functions. Remarkably, both the count of fundamental functions and the attributes connected to each one,

**Table 1** Variable unit and description

| Variables | Unit and description |
|---|---|
| Age | Years |
| Systolic blood pressure (SBP) | mmHg |
| Diastolic blood pressure (DBP) | mmHg |
| Waist-hip ratio (WHR) | waist circumference/hip circumference |
| Leukocyte (WBC) | $*10^3/\mu L$ |
| Hemoglobin (Hb) | $*10^6/\mu L$ |
| Platelets (Plt) | $*10^3/\mu L$ |
| Total bilirubin (TBIL) | mg/dL |
| Albumin (Alb) | mg/dL |
| Globulin (Glo) | g/dL |
| Alkaline Phosphatase (ALP) | IU/L |
| Serum glutamic oxaloacetic transaminase (GOT) | IU/L |
| Serum glutamic pyruvic transaminase (GPT) | IU/L |
| Serum γ-glutamyl transpeptidase (γ-GT) | IU/L |
| Lactate dehydrogenase (LDH) | mg/dL |
| Creatinine (Cr) | mg/dL |
| Uric acid (UA) | mg/dL |
| Triglycerides (TG) | mg/dL |
| High-density lipoprotein cholesterol (HDL-C) | mg/dL |
| Low-density lipoprotein cholesterol (LDL-C) | mg/dL |
| Plasma calcium concentration (Ca) | mg/dL |
| Plasma phosphate concentration (P) | mg/dL |
| Thyroid stimulating hormone (TSH) | μIU/mL |
| C reactive protein (CRP) | mg/dL |
| Education level (Edu.) | (1) Illiterate (2) Elementary school (3) Junior high school (4) High school (vocational) (5) Junior college (6) University (7) Graduate school or above |
| Marriage status (MS) | (1) Unmarried (2) Married |
| Income level (IL) | NTD/years<br>(1) Below $200,000 (2) $200,001–$400,000 (3) $400,001–$800,000 (4) $800,001–$1,200,000 (5) $1,200,001–$1,600,000 (6) $1,600,001–$2,000,000 (7) More than $2,000,000 |
| Drinking area | Alcohol proof × drinking years × frequency |
| Sport time | Hours per week × years of exercise × types |
| Sleeping hours (SH) | (1) 0~4 hours (2) 4~6 hours (3) 6~8 hours (4) more than 8 hours |
| insulin resistance (HOMA-IR) | FPI (μU/mL) × FPG (mg/dL)/405 |

encompassing product degree and knot placements, are autonomously established through data-driven mechanisms [21]. This strategy draws inspiration from the principles of recursive partitioning, akin to methods like Classification and Regression Trees in terms of its proficiency in capturing intricate higher-order interactions.

In the analysis phase, the dataset was initially partitioned into an 80% training dataset for use in model construction, and a separate 20% testing dataset for model assessment. In the training phase, MARS uses

**Table 2** *T*-test between insulin resistance and marital status, analysis of variance between insulin resistance and Income level, education level, and sleep hours

| | MS | IL | Edu. | SH |
|---|---|---|---|---|
| HOMA-IR | 0.996 | 0.001*** | 0.000*** | 0.000*** |

Note: MS: marriage status; IL: income level; Edu.: education level; SH: sleep hours; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

**Table 3** The *r* values of Pearson's correlation between insulin resistance and demographic, biochemistry, and lifestyle parameters

| | Age | SBP | DBP | WHR | WBC | Hb |
|---|---|---|---|---|---|---|
| HOMA-IR | 0.006 | 0.340*** | 0.27*** | 0.423*** | 0.335*** | 0.096*** |
| | Plt | TBIL | Alb | Glo | ALP | GOT |
| HOMA-IR | 0.217*** | –0.204*** | 0.057*** | 0.151*** | 0.245*** | 0.256*** |
| | GPT | γ-GT | LDH | Cr | UA | TG |
| HOMA-IR | 0.392*** | 0.215*** | 0.206*** | 0.001 | 0.310*** | 0.390*** |
| | HDL-C | LDL-C | Ca | P | TSH | CRP |
| HOMA-IR | –0.326*** | 0.220*** | 0.067*** | –0.096*** | 0.021* | 0.281*** |
| | Drinking area | | Sport time | | | |
| HOMA-IR | –0.045*** | | –0.084*** | | | |

Note: SBP: systolic blood pressure; DBP: diastolic blood pressure; WHR: waist-hip ratio; WBC: leukocyte; Hb: hemoglobin; Plt: platelets; TBIL: total bilirubin; Alb: albumin; Glo: globulin; ALP: alkaline phosphatase; GOT: serum glutamic oxaloacetic transaminase; GPT: serum glutamic pyruvic transaminase; γ-GT: serum γ-glutamyl transpeptidase; LDH: lactate dehydrogenase; Cr: creatinine; UA: uric acid; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; Ca: plasma calcium concentration; P: plasma phosphate concentration; TSH: thyroid stimulating hormone; CRP: C reactive protein;*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

specific hyperparameters that requires tuning to ensure good model performance. To facilitate this, the training dataset was once more divided at random to yield one segment for model formulation using a distinct set of hyperparameters, while the other section was used for validation purposes. A systematic exploration of all conceivable combinations of hyperparameters was performed using a comprehensive grid search approach. Subsequently, the model characterized by the lower root mean square error when applied to the validation dataset was deemed the optimal choice for each compared to MLR.

In the evaluation phase, the testing dataset was used to gauge the predictive efficacy of the MARS model. Given that the target variable in this study is a numerical parameter, the evaluation metrics chosen to compare model performance include symmetric mean absolute percentage error (SMAPE), relative absolute error (RAE), root relative squared error (RRSE), and root mean squared error (RMSE) (see Table 4).

To establish a comparative context, the averaged metrics derived from the MARS model were used to juxtapose the model performance with that of the benchmark MLR model. Note that both the MARS models and the MLR model were trained and tested using the same dataset.

In this study, all methods were performed using R software version 4.0.5 and RStudio version 1.1.453 with the required packages installed (http://www.R-project.org, accessed on 11 June 2024; https://www.rstudio.com/products/rstudio/, accessed on 11 June 2024). The implementations of MARS were the "earth" R package version 5.3.3 [22] , the "caret" R package version 6.0–94 [23]. MLR was implemented using the "stats" R package version 4.0.5, using the default settings to construct the models.

## Results

A total of 4,071 healthy subjects were enrolled. Demographic data are summarized in Table 5.

Table 6 shows the PCA result. The respective contributions of WBC and CRP are 0.62 and 0.38. Though CRP has a smaller impact on IR, it was not negligible.

Table 7 shows that MARS yielded smaller prediction errors than the MLR method, indicating greater accuracy.

**Table 4** Equations for calculating performance metrics

| Metric | Description | Calculation |
|---|---|---|
| SMAPE | Symmetric mean absolute Percentage error | $SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i - \hat{y}_i\rvert}{(\lvert y_i\rvert + \lvert \hat{y}_i\rvert)/2}$ |
| RAE | Relative absolute error | $RAE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i)^2}}$ |
| RRSE | Root relative squared error | $RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}}$ |
| RMSE | Root mean squared error | $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ |

$\hat{y}_i$ and $y_i$ represent predicted and actual values, respectively; $n$ stands for the number of instances.

**Table 5** Participant demographics and the testing conditions for insulin resistance and various subgroup variables

| Demographics Variables | Mean ± SD | Biochemistry Variables | Mean ± SD |
|---|---|---|---|
| Age | 36.68 ± 7.75 | Alb | 4.39 ± 0.21 |
| Edu. | *N* (%) | Glo | 3.16 ± 0.33 |
| Illiterate | 1 (0.02%) | ALP | 50.98 ± 19.91 |
| Elementary school | 8 (0.20%) | GOT | 19.98 ± 9.40 |
| Junior high school | 32 (0.79%) | GPT | 19.85 ± 16.43 |
| High school (vocational) | 584 (14.35%) | γ-GT | 20.21 ± 37.54 |
| Junior college | 620 (15.23%) | LDH | 148.6 ± 24.83 |
| University | 2,255 (55.39%) | Cr | 0.81 ± 0.09 |
| Graduate school or above | 571 (14.03%) | UA | 4.71 ± 1.03 |
| MS | *N* (%) | TG | 80.46 ± 57.01 |
| Single | 1,678 (41.22%) | HDL-C | 63.78 ± 14.29 |
| With spouse | 2,393 (58.78%) | LDL-C | 108.16 ± 30.52 |
| IL | *N* (%) | Ca | 9.51 ± 0.37 |
| Below $200,000 | 731 (17.96%) | P | 3.84 ± 0.45 |
| $200,001–$400,000 | 972 (23.88%) | TSH | 1.79 ± 1.50 |
| $400,001–$800,000 | 1,446 (35.52%) | CRP | 0.20 ± 0.29 |
| $800,001–$1,200,000 | 629 (15.45%) | HOMA-IR | 1.76 ± 1.23 |
| $1,200,001–$1,600,000 | 140 (3.44%) | Lifestyle Variables | Mean ± SD |
| $1,600,001–$2,000,000 | 70 (1.72%) | Drinking area | 1.36 ± 6.17 |
| More than $2,000,000 | 83 (2.04%) | Sport time | 3.84 ± 6.28 |
| Biochemistry Variables | Mean ± SD | SH | *N* (%) |
| SBP | 108.84 ± 13.76 | <4 hours/day | 40 (0.98%) |
| DBP | 70.84 ± 9.67 | 4–6 hours/day | 1,085 (26.65%) |
| WHR | 0.76 ± 0.05 | 6–7 hours/day | 1,841 (45.22%) |
| WBC | 6.04 ± 1.59 | 7–8 hours/day | 884 (21.71%) |
| Hb | 13.01 ± 1.35 | 8–9 hours/day | 177 (4.35%) |
| Plt | 251.14 ± 59.04 | >9 hours/day | 44 (1.08%) |
| TBIL | 0.89 ± 0.34 | | |

**Table 6** The results of principal component analysis of white blood cell count and c-reactive protein

| Component | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| WBC | 1.24889 | 0.49772 | 0.6244 | 0.6244 |
| CRP | 0.751114 | | 0.3756 | 1 |

WBC: leukocyte, CRP: c-reactive protein

**Table 7** The average performance of multiple linear regression and multivariate adaptive regression splines

| Methods | SMAPE | RAE | RRSE | RMSE |
|---|---|---|---|---|
| MARS | 0.501 | 1.225 | 1.150 | 1.694 |
| MLR | 0.522 | 1.256 | 1.179 | 1.738 |

Note: MLR: multiple linear regression, MARS: multivariate adaptive regression splines.

**Table 8** List of basis function $B_i$ of the MARS model and their coefficients, $a_i$

| | Definition | $a_i$ |
|---|---|---|
| Intercept | — | 6.634 |
| *B*1 | Max(0, 141 – SBP) | –0.013 |
| *B*2 | Max(0, 0.833 – WHR) | –1.448 |
| *B*3 | Max(0, WHR – 0.833) | 10.152 |
| *B*4 | Max(0, 11.53 – WBC) | –0.071 |
| *B*5 | Max(0, WBC – 11.53) | 0.681 |
| *B*6 | Max(0, 0.83 – TBIL) | 1.047 |
| *B*7 | Max(0, 24 – GOT) | 0.043 |
| *B*8 | Max(0, GOT – 24) | –0.017 |
| *B*9 | Max(0, 100 – GPT) | –0.033 |
| *B*10 | Max(0, GPT – 100) | 0.013 |
| *B*11 | Max(0, UA – 8.5) | 1.062 |
| *B*12 | Max(0, 303 – TG) | –0.004 |
| *B*13 | Max(0, 59 – HDL-C) | 0.021 |
| *B*14 | Max(0, HDL-C – 59) | –0.005 |
| *B*15 | Max(0, 0.7 – CRP) | –1.351 |
| *B*16 | Max(0, CRP – 0. 7) | –0.449 |

WHR: waist-hip ratio, SBP: systolic blood pressure, TG: triglyceride, HDL-C: high-density lipoprotein cholesterol, GOT: serum glutamic oxaloacetic transaminase, GPT: serum glutamic pyruvic transaminase, TBIL: total bilirubin, UA: uria acid, CRP: C-reactive protein

Table 8 shows the 16 basis functions (BF) derived from MARS from the following variables: SBP, WHR, WBC, TBIL, GOT, GPT, UA, TG, HDL-C, and CRP. Based on Table 8, the MARS equation is generated as follows: HOMA-IR = 6.634 – 1.448MAX(0, 0.833 – WHR) + 10.152MAX(0, WHR – 0.833) – 1.351MAX(0, 0.7 – CRP) – 0.449MAX(0, CRP – 0.7) + 1.062MAX(0, UA – 8.5) + +1.047(MAX(0, 0.83 – TBIL) + 0.681MAX(0, WBC – 11.53) – 0.071MAX(0, 11.53 – WBC) + 0.043MAX(0, 24 – GOT) – 0.017MAX(0, GOT – 24) + 0.021MAX(0, 59 – HDL) – 0.005MAX(0, HDL – 59) – 0.013MAX(0, 141 – SBP) – 0.033MAX(0, 100 – GPT) + 0.013MAX(0, GPT – 100) – 0.004MAX(303 – TG)

These BFs are interpreted as follows. Taking WHR as an example, when the WHR was below 0.833, the first equation was used: HOMA-IR = –1.448 × (0.833 – WHR). When subject WHR exceeded 0.8333, the equation changed to HOMA-IR = 10.152 × (WHR – 0.833333). In short, the results of the aforementioned equations should all exceed 0. For those risk factors with more than one BF, corresponding figures are displayed to capture a clearer understanding of the relationship between these factors and HOMA-IR. An equation built with these BFs is shown in Fig. 2.

Finally, the overview of the present study is shown in Graphical Abstract. It could be noted that, from how to selecting the participants, to the Mach-L methods we used, to the comparison between these two methods and, the equation built by MARS are all depicted.

## Discussion

The present study constructed an equation to estimate HOMA-IR in healthy Chinese women aged between 20–50. This age range was selected to exclude menopausal women. No subjects were currently using medications known to affect blood pressure (BP), FPG or blood lipids, and had no major diseases. Our results found that there were 10 biochemistry variables related to HOMA-IR. To our knowledge, the present study is the only one using MARS to build the equation. It is of clinical use and, in the same time, could shed light on the true determining factors for IR in premenopausal women.

No lifestyle factors were found to have significant correlation with HOMA-IR, including smoking, drinking, marital status, income, educational attainment, and exercise. This does not necessarily indicate that these factors are not important for insulin resistance, but rather suggest that their impact might be “absorbed” by their end results such as higher BP or WHR. A similar explanation could also be applied to the role of age. As

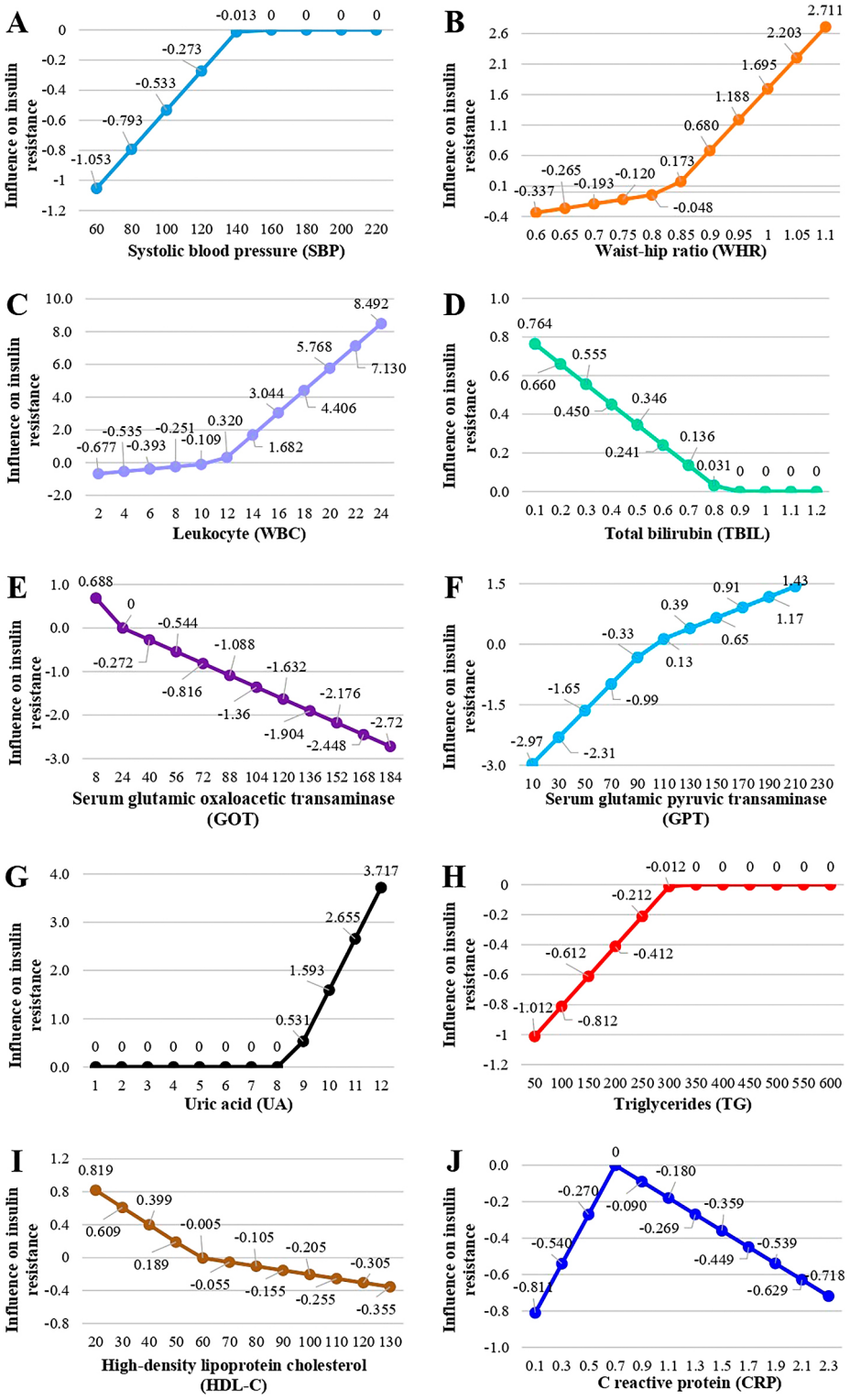
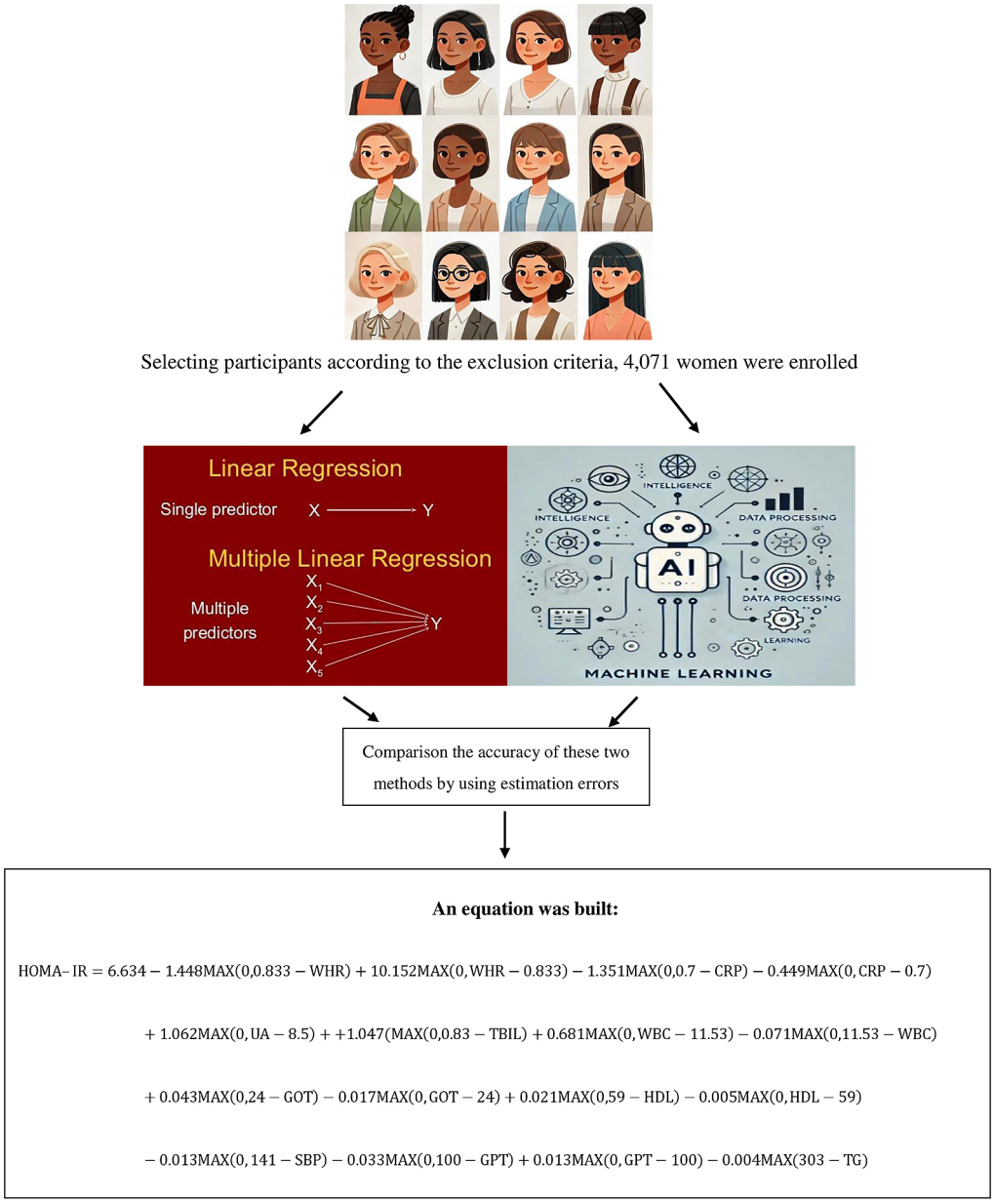



**Fig. 2** Influence of important variables on the insulin resistance

A: Systolic blood pressure. B: Waist-hip ratio. C: Leukocyte. D: Total bilirubin. E: Serum glutamic oxaloacetic transaminase. F: Serum glutamic pyruvic transaminase. G: Uric acid. H: Triglycerides. I: High-density lipoprotein cholesterol. J: C reactive protein.

mentioned in the introduction section, the impact of age on IR remains controversial. Our results found no correlation, suggesting that the positive relationship found in other studies might be due to differences in study design, age groupings, and ethnicity. IR might also be indirectly related to age through factors such as inflammation or body weight.

Ten biochemistry variables were found to be significantly




**Graphical Abstract** Overview of the study methodology, including participant selection criteria, the Mach-L methods applied, the comparative analysis between the two methods, and the equation developed using MARS.

related to HOMA-IR, the six most important of which are discussed here. WHR had the most significant impact. Many previous studies examined the relationship between obesity and IR [24]. Obesity can be measured using many methods, including body fat, body mass index, and WHR. Body fat measurement is more expensive and requires specific hardware, raising clinical obstacles. While simpler to use, body mass index cannot determine whether body weight is due to fat, muscle, or bone mass, making it less accurate [25]. Previous studies have shown that WHR provides greater accuracy than body mass index [26]. The results of the present study suggest that, among the ten biochemistry variables examined, WHR has the most significant impact on HOMA-IR. A review article published by Wondmkun suggested that this relationship was caused by multiple mechanisms, such as increased amounts of non-esterified fatty acids, glycerol, hormones, and pro-inflammatory cytokines [27].

The second leading factor was UA. In an empirical study of 687 diabetic patients, Han *et al.* found that UA is positively related to metabolic score for IR (METS-IR) ($r$ = 0.238, $p$ < 0.01) [28]. While that study focused on diabetics rather than healthy individuals, their results can be seen as supporting the findings of the present study. This raises the possibility of improving IR by reducing

UA. Takir *et al.* found that HOMA-IR decreased after allopurinol treatment ($n$ = 73) when compared to controls ($n$ = 33) in a group of asymptomatic hyperuricemia subjects [29]. UA could increase HOMA-IR through several pathways such as inhibition of Adenosine 5'-monophosphate (AMP)-activated protein kinase, thereby increasing gluconeogenesis, inflammation, and oxidative stress in the islet cells [30-33].

As a systemic inflammatory biomarker, CRP is regarded as a strong and independent predictor for cardiovascular diseases, diabetes, and hypertension [34]. It reflects only local inflammation in atherosclerosis, but is also linked to IR [35]. Our results found CRP to be third leading factor, which is consistent with other mainstream studies. For example, Gelaye *et al.* showed that women with higher CRP had a 2.18-fold increased risk of HOMA-IR compared to the lowest tertile. In men, the heightened risk increased to 2.54-fold. This relationship could be explained by CRP being synthesized in liver and regulated by interleukin-6 and tumor necrosis factor-alfa [36]. At the same time, FPG and FPI values would increase in the presence of chronic and systemic inflammation [35].

Surprisingly, bilirubin was the next key factor. The few studies that have examined the relationship of bilirubin to IR have found a negative correlation [37, 38]. Working with a small sample, Zhang *et al.* reported that indirect bilirubin might reduce IR. Takei *et al.* suggested that this correlation might be related to bilirubin's unique ability to suppress cytokines [39]. Our findings strongly support this correlation in a large cohort.

The correlation between WBC and IR remains controversial. Mahdiani *et al.* found no relationship between WBC and HOMA-IR in 283 diabetic patients [40], while Kuo *et al.* found a positive correlation in 21,112 non-obese men, though they used an equation based on demographic and biochemistry data to estimate IR which differs from HOMA-IR [41]. Mahdiani's results could be explained by diabetic patients typically having high IR values, which may mask the relationship. Our findings provide additional evidence that, although both WBC and CRP are markers for inflammation, they have independent influences on HOMA-IR.

Several reports have shown that diabetes and metabolic syndrome are related to liver enzymes [42] based on a correlation between liver enzymes and IR. In a study of 261 subjects, Niranjan *et al.* found that the crude prevalence ratio for higher HOMA-IR (≥3.8) was increased in subjects with elevated GOT and GPT. However, the significance for GOT disappeared in multivariate analysis [43]. Our results are consistent with these findings, but both GOT and GPT had independent effects on HOMA-IR. This discrepancy could be explained by differences in study design, sample ethnicity, and study cohort.

The remaining three factors HDL-C, SBP and TG all correlated with IR [44-46], but given their relatively low importance, discussion is omitted here in consideration of article length.

The present study is subject to certain limitations. First, this is a cross-sectional study and is thus less persuasive than a longitudinal one. Second, only Chinese women were enrolled thus caution should be taken in extrapolating the findings to other ethnic groups.

## Conclusion

MARS is used to build an equation to estimate IR in Chinese premenopausal women with WHR, CPR, UA, TBIL, WBC, GOT, HDL, SPB, GPT, TG as following.

HOMA-IR = 6.634 – 1.448MAX(0, 0.833–WHR) + 10.152MAX(0, WHR – 0.833) – 1.351MAX(0, 0.7 – CRP) – 0.449MAX(0, CRP – 0.7) + 1.062MAX(0, UA – 8.5) + +1.047(MAX(0, 0.83 – TBIL) + 0.681MAX(0, WBC – 11.53) – 0.071MAX(0, 11.53 – WBC) + 0.043MAX(0, 24 – GOT) – 0.017MAX(0, GOT – 24) + 0.021MAX(0, 59 – HDL) – 0.005MAX(0, HDL – 59) – 0.013MAX(0, 141 – SBP) – 0.033MAX(0, 100 – GPT) + 0.013MAX(0, GPT – 100) – 0.004MAX(303 – TG)

This provides a precise and easily implemented tool when FPI level measurements are unavailable.

## Declaration

### *Ethical approval and consent to participate*

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Kaohsiung Armed Forces General Hospital (protocol code KAFGHIRB 112-006 and date of approval 21/06/2023).

### *Consent for publication*

Not applicable. Analysis was based on secondary data sourced from the MJ Health Research Foundation.

### *Availability of data and materials*

Data are available on request due to privacy/ethical restrictions.

### *Competing interests*

The authors declare no conflicts of interest.

### *Funding*

This research was funded by Kaohsiung Armed Forces General Hospital, grant number KAFGH_E_112053.

***Acknowledgements***

The authors thank all subjects who participated in the study.

***Author Contributions***

Validation: Dee Pei and Shih-Peng Mao; Formal analysis: Dee Pei; Investigation: Chen-Yu Wang, Chi-Hao Liu, Chung-Bao Hsieh, and Ta-Wei Chu; Data curation: Shih-Peng, Mao; Writing – original draft: Shih-Peng, Mao; Writing – review & editing: Yao-Jen Liang.

## References

1. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, *et al.* (2020) Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. *J Epidemiol Glob Health* 10: 107–111.
2. (2021) IDF Diabetes Atlas 10th edition 2021. Taiwan diabetes report 2000–2045. https://diabetesatlas.org/data/en/country/194/tw.html accessed on November 8, 2021.
3. (1999) Tseng CH. The costs of diabetes. http://www.tsim.org.tw/journal/jour10-6/P10_217.PDF accessed on December, 1999.
4. Galicia-Garcia U, Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, *et al.* (2020) Pathophysiology of type 2 diabetes mellitus. *Int J Mol Sci* 21: 6275.
5. Lebovitz HE (2001) Insulin resistance: definition and consequences. *Exp Clin Endocrinol Diabetes* 109 Suppl 2: S135–S148.
6. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, *et al.* (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28: 412–419.
7. Facchini FS, Hua N, Abbasi F, Reaven GM (2001) Insulin resistance as a predictor of age-related diseases. *J Clin Endocrinol Metab* 86: 3574–3578.
8. Barbieri M, Ragno E, Benvenuti E, Zito GA, Corsi A, *et al.* (2001) New aspects of the insulin resistance syndrome: impact on haematological parameters. *Diabetologia* 44: 1232–1237.
9. Tzou SJ, Peng CH, Huang LY, Chen FY, Kuo CH, *et al.* (2023) Comparison between linear regression and four different machine learning methods in selecting risk factors for osteoporosis in a Chinese female aged cohort. *J Chin Med Assoc* 86: 1028–1036.
10. Wu X, Tsai SP, Tsao CK, Chiu ML, Tsai MK, *et al.* (2017) Cohort profile: The Taiwan MJ cohort: half a million Chinese with repeated health surveillance data. *Int J Epidemiol* 46: 1744–1744g.
11. Foundation MHR. The introduction of MJ health database. MJ Health Research Foundation Technical Report, MJHRF-TR-01. 2016 http://www.mjhrf.org/file/file/report/MJHRF-TR-01%20MJ%20Health%20Database.pdf accessed on August 22, 2016.
12. Deconinck E, Zhang MH, Petitet F, Dubus E, Ijjaali I, *et al.* (2008) Boosted regression trees, multivariate adaptive regression splines and their two-step combinations with multiple linear regression or partial least squares to predict blood-brain barrier passage: a case study. *Anal Chim Acta* 609: 13–23.
13. Rashijane LT, Mokoena K, Tyasi TL (2023) Using multivariate adaptive regression splines to estimate the body weight of savanna goats. *Animals (Basel)* 13: 1146.
14. Toutounji H, Durstewitz D (2018) Detecting multiple change points using adaptive regression splines with application to neural recordings. *Front Neuroinform* 12: 67.
15. Jalali-Heravi M, Asadollahi-Baboli M, Mani-Varnosfaderani A (2009) Shuffling multivariate adaptive regression splines and adaptive neuro-fuzzy inference system as tools for QSAR study of SARS inhibitors. *J Pharm Biomed Anal* 50: 853–860.
16. Pfob A, Lu SC, Sidey-Gibbons C (2022) Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Med Res Methodol* 22: 282.
17. Gackowski M, Szewczyk-Golec K, Pluskota R, Koba M, Mądra-Gackowska K, *et al.* (2022) Application of Multivariate Adaptive Regression Splines (MARSplines) for predicting antitumor activity of anthrapyrazole derivatives. *Int J Mol Sci* 23: 5132.
18. Menon R, Bhat G, Saade GR, Spratt H (2014) Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. *Acta Obstet Gynecol Scand* 93: 382–391.
19. Peters-Sanders L, Sanders H, Goldstein H, Ramachandran K (2023) Using multivariate adaptive regression splines to predict lexical characteristics' influence on word learning in first through third graders. *J Speech Lang Hear Res* 66: 589–604.
20. Zakeri IF, Adolph AL, Puyau MR, Vohra FA, Butte NF (2010) Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. *J Appl Physiol (1985)* 108: 128–136.
21. Friedman JH, Roosen CB (1995) An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 4: 197–217.
22. (2024) Milborrow S. Derived from Mda: MARS by T. Hastie and R. Tibshirani. Earth: Multivariate Adaptive Regression Splines. R Package Version, 5.3.3. http://CRAN.R-project.org/package=earth accessed on February 26, 2024.
23. (2023) Kuhn M. Caret: Classification and Regression Training. R Package Version, 6.0–94. 2023. https://

CRAN.R-project.org/package=caret accessed on March 21, 2023.
24. Sirbu AE, Buburuzan L, Kevorkian S, Martin S, Barbu C, *et al.* (2018) Adiponectin expression in visceral adiposity is an important determinant of insulin resistance in morbid obesity. *Endokrynol Pol* 69: 252–258.
25. (2011) Centers for Disease Control and Prevention (U.S.). Body mass index: considerations for practitioners. Available online: https://stacks.cdc.gov/view/cdc/25368 accessed on August 2: 2011.
26. (2023) Pugle M. BMI Is Outdated—Here's Why Your Waist-to-Hip Ratio Is a Better Indicator of Health. https://www.health.com/waist-to-hip-ratio-health-indicator-8285898 accessed on October 6, 2023.
27. Wondmkun YT (2020) Obesity, insulin resistance, and type 2 diabetes: associations and therapeutic implications. *Diabetes Metab Syndr Obes* 13: 3611–3616.
28. Han R, Zhang Y, Jiang X (2022) Relationship between four non-insulin-based indexes of insulin resistance and serum uric acid in patients with type 2 diabetes: a cross-sectional study. *Diabetes Metab Syndr Obes* 15: 1461–1471.
29. Takir M, Kostek O, Ozkok A, Elcioglu OC, Bakan A, *et al.* (2015) Lowering uric acid with allopurinol improves insulin resistance and systemic inflammation in asymptomatic hyperuricemia. *J Investig Med* 63: 924–929.
30. Lanaspa MA, Cicerchi C, Garcia G, Li N, Roncal-Jimenez CA, *et al.* (2012) Counteracting roles of AMP deaminase and AMP kinase in the development of fatty liver. *PLoS One* 7: e48801.
31. Lanaspa MA, Sanchez-Lozada LG, Choi YJ, Cicerchi C, Kanbay M, *et al.* (2012) Uric acid induces hepatic steatosis by generation of mitochondrial oxidative stress: potential role in fructose-dependent and -independent fatty liver. *J Biol Chem* 287: 40732–40744.
32. Cicerchi C, Li N, Kratzer J, Garcia G, Roncal-Jimenez CA, *et al.* (2014) Uric acid-dependent inhibition of AMP kinase induces hepatic glucose production in diabetes and starvation: evolutionary implications of the uricase loss in hominids. *Faseb J* 28: 3339–3350.
33. Roncal-Jimenez CA, Lanaspa MA, Rivard CJ, Nakagawa T, Sanchez-Lozada LG, *et al.* (2011) Sucrose induces fatty liver and pancreatic inflammation in male breeder rats independent of excess energy intake. *Metabolism* 60: 1259–1270.
34. Gelaye B, Revilla L, Lopez T, Suarez L, Sanchez SE, *et al.* (2010) Association between insulin resistance and c-reactive protein among Peruvian adults. *Diabetol Metab Syndr* 2: 30.
35. Nakanishi N, Shiraishi T, Wada M (2005) Association between C-reactive protein and insulin resistance in a Japanese population: the Minoh Study. *Intern Med* 44: 542–547.
36. Gabay C, Kushner I (1999) Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med* 340: 448–454.
37. Guzek M, Jakubowski Z, Bandosz P, Wyrzykowski B, Smoczyński M, *et al.* (2012) Inverse association of serum bilirubin with metabolic syndrome and insulin resistance in Polish population. *Przegl Epidemiol* 66: 495–501.
38. Zhang F, Guan W, Fu Z, Zhou L, Guo W, *et al.* (2020) Relationship between serum indirect bilirubin level and insulin sensitivity: results from two independent cohorts of obese patients with impaired glucose regulation and type 2 diabetes mellitus in China. *Int J Endocrinol* 2020: 5681296.
39. Takei R, Inoue T, Sonoda N, Kohjima M, Okamoto M, *et al.* (2019) Bilirubin reduces visceral obesity and insulin resistance by suppression of inflammatory cytokines. *PLoS One* 14: e0223302.
40. Mahdiani A, Kheirandish M, Bonakdaran S (2019) Correlation between white blood cell count and insulin resistance in type 2 diabetes. *Curr Diabetes Rev* 15: 62–66.
41. Kuo TY, Wu CZ, Lu CH, Lin JD, Liang YJ, *et al.* (2020) Relationships between white blood cell count and insulin resistance, glucose effectiveness, and first- and second-phase insulin secretion in young adults. *Medicine (Baltimore)* 99: e22215.
42. Music M, Dervisevic A, Pepic E, Lepara O, Fajkic A, *et al.* (2015) Metabolic syndrome and serum liver enzymes level at patients with type 2 diabetes mellitus. *Med Arch* 69: 251–255.
43. Niranjan S, Phillips BE, Giannoukakis N (2023) Uncoupling hepatic insulin resistance - hepatic inflammation to improve insulin sensitivity and to prevent impaired metabolism-associated fatty liver disease in type 2 diabetes. *Front Endocrinol (Lausanne)* 14: 1193373.
44. Iwani NA, Jalaludin MY, Zin RM, Fuziah MZ, Hong JY, *et al.* (2017) Triglyceride to HDL-C ratio is associated with insulin resistance in overweight and obese children. *Sci Rep* 7: 40055.
45. Zhou MS, Wang A, Yu H (2014) Link between insulin resistance and hypertension: what is the evidence from evolutionary biology? *Diabetol Metab Syndr* 6: 12.
46. Berthezene F (1992) Hypertriglyceridemia: cause or consequence of insulin resistance? *Horm Res* 38: 39–40.