Original Research   **Open Access**

Examining the association between C-Reactive protein and obesity by using the fractional polynomial approach; applying on NHANES dataset from 2001 to 2010

Ghada Abo-Zaid[1,2*] and Karyn Morrissey[1]

*Correspondence: K.Morrissey@exeter.ac.uk (or) abozaidg@gmail.com

CrossMark
← Click for updates

[1]European Centre for Environment and Human Health, University of Exeter Medical School, Knowledge Spa, Royal Cornwall Hospital, Truro, Cornwall TR1 3HD United Kingdom.
[2]Faculty of Commerce- Ain Shams University, Khalifa El-Maamon St, Abbaisiya Sq., Cairo 11566, Egypt.

Abstract

Objective: This study uses a flexible nonlinear approach, Fractional polynomial models (FPs), to examine the association between obesity and C-reactive protein to select the best fitted model within 44 potentially FP models.

Methods: Data for 5 years (2001-2010) of the National Health Interview Survey (NHANES) was used. All respondents aged between 17 and 74 were included in the analysis. CRP was transformed to ln(CRP) to eliminate skewness and missing values were removed from the analysis. A fractional polynomial approach was applied to measure the relationship between elevated levels of CRP and obesity. A closed test was used to select the best model among the 44 models.

Results: The best fitted fractional polynomial regression model contained the powers -2 and -2 for BMI. The association between the ln(CRP) and BMI when estimated using the FP approach exhibited a J-shaped pattern for women and men. Women have a higher risk of elevated CRP level compared to men. A deviance difference test yielded a significant improvement in model fit of -2 and -2 compared to other BMI functions.

Conclusion: The fractional polynomial regression model is the most robust estimator of BMI compared to other linear or nonlinear models.

Keywords: Categorization, C-reactive protein, fractional polynomial model, linear model, obesity

Introduction

C-reactive protein (CRP) is an acute-phase protein of the family of the pentraxins and is widely used in clinical settings to monitor chronic and acute inflammatory conditions. Recent research has found that the increase of BMI is associated with elevated CRP concentrations regardless of sex, age, and ethnicity [1,2].

Various models are available to analyze the relationship between CRP and BMI. However, the choice between linear and non-linear analysis is controversial in applied fields such as medicine, clinical trials, and epidemiology. A few studies show that categorizing the continuous data is preferable , especially if the association between two variables is nonlinear. (Wang et al., 2016). However, Royston, et al(2006) [3]. Reported many pitfalls for this approach such as the loss of information and decrease the power of the model [3-7], while debate over the appropriate cut-off points for normal, overweight and obese further complicate the categorization of BMI.

A number of studies have used BMI as a continuous variable. However, estimating the association between BMI and a set of covariates using a continuous scale is challenging because the relationship may be nonlinear, and the BMI distribution is often right skewed. Furthermore, linear models require many assumptions including data normality and the absence of multicollinearity, and heteroscedasticity in the data. In addition, linear analysis assumes a constant influence of the independent variable on the outcome [5].

Previous research has used quadratic or cubic polynomials (non-linear approach), but the range of curve shapes afforded by low ordered polynomials is restricted [8]. In light of these

modeling limitations, the purpose of this study is to investigate a flexible approach to modeling the relationship between obesity and CRP. Combining the strengths of linear and non-linear models, FP models use polynomial transformations to measure the association between the independent and the dependent variables [4]. Fractional polynomials are more powerful than the regular polynomials and provide flexible transformation for continuous variables to determine the best fitting function form for BMI , by using the closed test [7,9,10].

The objectives of this paper are twofold. First this paper examines the association between CRP and obesity, and whether an association remains after adjusting for variables such as age, cotinine level (smoking status), alcohol consumption, race and gender. Second, we focus on comparing model outcomes between linear, nonlinear, categorization and multivariate fractional polynomial model (MFP) approaches to select the best-fitted model for the association between ln (CRP) level and BMI.

Methods
Study population
This study used National Health Interview Survey (NHANES) which is designed to examine the health and diet of non-institutionalized and civilian (children and adults) in the United States. The National Centre of Health Statistics (NCHS) at the Centre for Disease Control and Prevention [11] was responsible for setting up the NHANES surveys and its program of studies. NCHS has also approved the study protocol of the NHANES survey; for more details about the methods of the ethics approval, the design of the survey and the methodology study are available elsewhere [12,13]. The focus of this study was participants aged between 17 and 74 years old. Individuals were selected based on the availability of their CRP and their BMI measurements. The data for this study was based on five cross-sectional independent studies conducted from NHANS dataset, starting from 2001 to 2010. Later studies were not included because CRP was not measured post 2010. Pooling the data across years yielded a sample of 52,749 participants with 25,976 (44%) and 26,773(51%) men and women participants respectively.

Measurements, Laboratory variables, and other variables
Measurements
CRP: "The latex-enhanced nephelometry with high sensitivity by using a Dade Behring Nephelometer II Analyzer System (Dade Behring Diagnostics,Inc., Somerville, New Jersey) was used to measure CRP levels [13]. CRP level was measured on a continuous scale. For the purpose of this study it was transformed to ln(CRP) to eliminate skewness.

Laboratory variables
The medical examination centre assessment was responsible for measuring the laboratory variables. "Serum cotinine was measured by an isotope dilution–high-performance liquid

chromatography-atmospheric pressure chemical ionization-tandem mass spectrometry. Cotinine concentrations were derived from the ratio of native/labeled cotinine in the sample, by comparisons to a standard curve. Descriptions, in details, of serum cotinine measurement NHANES, are available on-line [14]. The cotinine level was used as an indicator of the participants' smoking status. It is used as a continuous scale on the statistical analyses for all models.

Other variables
Self-reported questionnaires were used for race, gender, and age at the baseline.

Race, in the NHANES dataset, was presented in five categories listed as American, Non-Hispanic white, non-Hispanic black, other Hispanic, and other races. Gender was classified into male and female while age recorded on a continuous scale. BMI is measured on a continuous scale, and it is calculated by weight (kg)/height (m²). A respondent was considered to be obese if their BMI was greater than 30. The individuals with (BMI >50 or BMI<18) were discarded from the analyses, to avoid the outliers.

Data analysis
Five cross-sectional independent studies from NHANES dataset were used using data for 2001 to 2010. This data pooled together yielded a sample of 52749 participants with 25967 men and 26773 women, respectively. Weighting variables were used to account for the complex sample design of the NHANES. Each of the statistical analyses were stratified by gender to consider the differential biological factors that lead to weight gain between men and women [12]. The association between CRP level and BMI was modeled using linear, non-linear, categorical, and FP statistical approaches. Each statistical model was adjusted for age, cotinine level (smoking history) and alcohol consumption.

Statistical models
Categorical model
In this study, we categorize BMI into three groups (obesity: BMI>30, overweight: 30>BMI>25, and normal: 18.5<BMI<25) based on World health organization (WHO) criteria. The (unadjusted) categorization model can be written as follows:

$$\ln(CRP) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \varepsilon$$

Where X_1 and X_2 are BMI levels for overweight and obese participants. β_0 is constant while β_1, and β_2 are the influence of the overweight and obese individuals compared to the normal weight participants. *ln (CRP)* is the logarithm of CRP level, and is the error term.

Linear model
The linear regression (unadjusted) model can be written as follows:

$$\ln(CRP) = \beta_0 + \beta_3 X_{BMI} + \varepsilon$$

Where X_{BMI} is the BMI level on a continuous scale, and β_3 is the impact of increasing the BMI level.

Fractional polynomial approach

The fractional polynomial approach is a flexible model that combines aspects of the linear and nonlinear models. Essentially, the mechanism of this model depends on using polynomial transformations to estimate the relation between CRP level (outcome) and BMI level (covariate). Fractional polynomial generate 44 models; The first degree polynomial models (FP1) is based on one polynomial term and it estimates 8 models derived from the power of the following set {-2, -1, -0.5, 0, 0.5, 1, 2, 3}; if the power equals zero this means taking the logarithm of the BMI covariate (log(X))]. If the power term is one for BMI, this means the model estimate is linear. The FP1 model (unadjusted model) can be written as:

$$\ln\left(CRP\right) = \beta_0 + \beta_4 X_{BMI}^{P_1} + \varepsilon$$

Where P_1 is the first-degree power of the Fractional polynomial model. It transforms the BMI level based on the set {-2, -1, -0.5, 0, 0.5, 1, 2, 3} and produced 8 possible FP1 models.

The second-degree fractional polynomial models (FP2) are based on two polynomials (power terms) and it estimates 36 model based on the same power set noted above [9]. The FP2 model (unadjusted model) can be written as

$$\ln\left(CRP\right) = \beta_0 + \beta_4 X_{BMI}^{P_1} + \beta_5 X_{BMI}^{P_2} + \varepsilon$$

Where p_1 and p_2 are the powers terms of FP2, when $p_1 = p_2$, eight models are expected with the same powers. The model will be on the form

$$\beta_4 X_{BMI}^{P_1} + \beta_5 X_{BMI}^{P_2} \log\left(X_{BMI}\right)$$

only when the powers are the same. The closed test is used to select the best-fitted model among the 44 models [9].

Assessment criteria

To assess the strength of the fractional polynomial approach for BMI compared to the other three approaches (e.g., linear, polynomial and categorization model), the models were stratified by gender. The assessment was repeated after adjusting the models by age, cotinine level (smoking history) and alcohol consumption. The models are compared using three methods; (i) Deviation difference, (ii) graphically based on the shape of the CRP level and BMI, (iii) Root mean square error (Root MSE). STATA statistical software (Version 12) was used to undertake each of the statistical analyses.

Results

Summary statistics of the NHANES dataset sample are shown in **Table 1**. In this study, 52749 participants were included from merging five continuous studies (2001 to 2010); this sample is classified to 26773 men and 25976 women with an aver-

age age of 44.89 and 44.25 years old respectively. Race and gender were measured on a categorical scale, while age and BMI were measured on a continuous scale (mean SD (**Table 1**). The majority of the respondents in the sample were in the overweight and obese classification; the percentage of obese in women is greater than men with 31.32 % and 26.01% respectively; while the percentage of overweight is more in men compared to women with 34.62% and 27.80 respectively. The percentage of people in the normal weight range is almost the same for both gender, Heavy smokers made up 27.85% of the male sample, and 26.22% of the female sample. The majority of the participants were non-smokers with 59% and 65% for men and women respectively. Ever smokers made up 54.93% of the male sample and 40.47% of the female sample. With regard to race, 42% of the sample was non-Hispanic white, followed by 25% Mexican-American and 25% non-Hispanic black for both men, and women samples, see **Table 1**.

Model fit and findings

The best-fitted model for obtained from the second-degree fractional polynomial model with powers ($P_1=-2$ and $P_2=-2$) for both male and female sample, where $\overline{BMI}$=BMI/100. After adjusting the model by age, cotinine level and alcohol consumption the models remain unchanging.   The findings of FP2 model were better than the FP1 model for both men and women with (Male: Deviance Difference=61.72, P-value <0.0001), and (Female: Deviance Difference=128.89, P-value<0.0001) respectively. The BMI- ln(CRP) curves for FP1 and FP2 models is illustrated in **Figure 1**. The FP2 model yielded a J-curve for the association between the BMI and CRP level for both gender; For instance, if BMI measure is less than 18, the impact on the CRP level is barely visible; while if BMI is over 18, the CRP level is slightly raised and it has significantly increased if BMI was over 30. This shows a positive direct association between CRP level and BMI for men and women
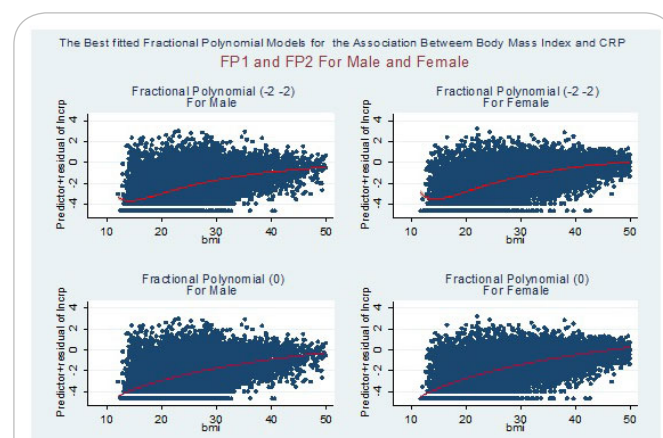

Figure 1. The Prediction Findings for the Association Between CRP Level and BMI for Male and Female Samples by Using the Best-Fitted FP2 Model With Powers ($p_1=-2$, $p_2=-2$) And FP1 Model With Power ($p_1=0$).

Table 1. Descriptive Statistics for NHANES rid Study From 2001 To 2010 Used to Estimate the Relationship Between the Elevated Level Of CRP and Obesity.

Survey Year	2001- 2002	2003 -2004	2005 -2006	2007-2008	2009-2010	Total
Male						
Sample Size	5605	4970	5080	5096	5225	25976
CRP : Mean (SD)	0.31()0.53	0.29(0.82)	0.41(0.83)	0.34(0.81)	0.30(0.73)	0.31
Age: Mean (SD)	40.7(17.57)	40.83(18.12)	40.52(17.48)	44.15(16.65)	43.53(16.85)	44.89
BMI Level (% Prevalence)						
Normal:(18.5 to25)	42.83	42.56	39.90	35.99	35.59	39.37
Overweight:(25-30)	35.27	33.74	34.00	35.50	34.57	34.62
Obese :(30)	21.91	23.7	26.1	28.51	29.83	26.01
BMI Level:Mean (SD)	24.51(6.33)	24.58(6.31)	24.58	25.21	25.38	24.86
Cotinine level: Mean (SD)	45.25	48.49	47.65	57	49.01	49.45(118.5)
Race(% Prevalence)						
Mexican American	24.61	24.87	27.15	21.11	22.85	24.11
Other Hispanic	4.57	3.34	3.21	11.11	10.43	6.53
Non-Hispanic White	41.96	40.36	38.54	41.56	42.14	40.95
Non-Hispanic Black	24.96	26.54	26.77	21.47	18.60	23.65
Other Races	3.91	4.89	4.33	4.75	5.97	4.76
Female						
Sample Size	5988	5152	5268	5053	5312	26773
CRP : Mean (SD)	0.58(0.75)	0.41(0.83)	0.42(0.83)	0.43(0.81)	0.38(0.64)	0.42
Age: Mean (SD)	39(17.15)	40.28(17.64)	38.78(17.07)	44.65(16.65)	43.54(16.55)	44.25
BMI Level (% Prevalence)						
Normal:(18.5 to25)	43.45	43.97	42.49	37.41	37.10	40.88
Overweight:(25-30)	28.79	27.08	26.55	28.60	28.00	27.80
Obese I: (30 -35)	27.76	28.95	30.95	33.99	34.9	31.33
BMI Level:Mean (SD)	25.07(7.22)	25.21(7.15)	25.16(7.41)	25.80(7.54)	25.98(7.63)	25.44(7.74)
Race Race(% Prevalence)						
Mexican American	24.79	24.90	27.87	21.39	22.40	24.30
Other Hispanic	4.79	3.40	3.53	12.57	11.07	6.99
Non-Hispanic White	42.00	41.28	37.40	39.52	41.75	40.44
Non-Hispanic Black	23.96	26.09	25.63	22.11	18.54	23.27
Other Races	4.46	4.33	5.58	4.41	6.23	5.00

in the sample, and the findings of the FP2 model were more accurate over FP1 with $(p_1=log(BMI))$.

In addition, the FP2 model produced more accurate results for the both the male and female sub-samples compared to the linear regression model (Male: Deviance difference=231.31, Female: Deviance difference=378.58 and P-value=0.0001), linear quadratic model (Male: Deviance difference=670.41, Female: Deviance difference=1074.65 and P-value=0.0001) and categorization model (Male: Deviance difference=2776.958, Female: Deviance difference=3011.603 and P-value=0.0001) respectively. **Figure 2** gives the predicted values for ln(CRP) with BMI for male and female sub-samples separately using the FP1 and FP2 models. The BMI-ln(CRP) curve produced J-shape and it was more precise than the FP1 model and it also become noticeable that the risk of elevated level of CRP level is more in women over men.

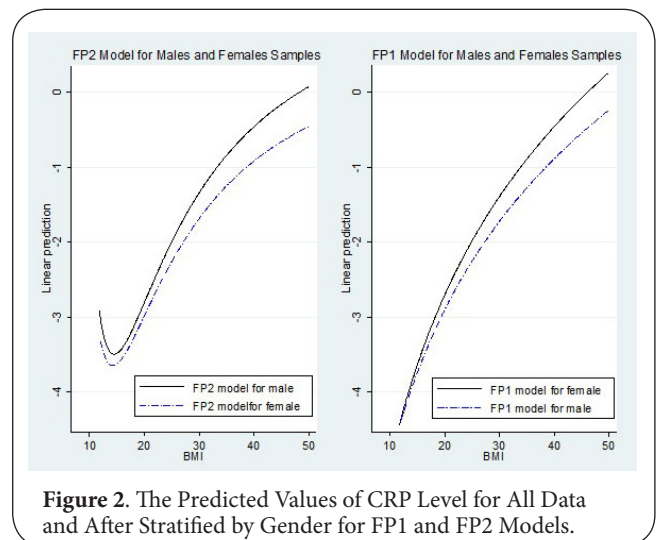Using the FP2 approach, the male and female sub-samples

Figure 2. The Predicted Values of CRP Level for All Data and After Stratified by Gender for FP1 and FP2 Models.

were further stratified by race (Mexican American, other Hispanic, non-Hispanic white, non-Hispanic black, and other races), and produced five ln(CRP)-BMI curves, see **Figure 3**. In the male sub-sample, the best FP2 model was the same with powers *(p₁=-2, p₂=-2)* for all race groups except the 'other Hispanic', and 'others groups' with powers *(p₁=-1, p₂=-1)*, and $p_1=2, p_2=2)$ When BMI is greater than 30, the level of CRP was slightly higher for the other-Hispanic group followed by non-Hispanic black compared to others. In the female sub-sample, all race groups produced the same FP2 models with powers *(p₁=-2, p₂=-2)*, and yielded a J-shape. The Mexican American group had the highest increase in CRP level and the and the other races group the lowest, see **Figure 3**.

Figure 4 presents the fitted curves of the association between ln(CRP) level and BMI at the 95% confidence interval for the FP2 model. The curves were for men and women who are never smoke at age 40, 50 and 60 respectively, with a BMI greater than 18.5. The estimated association between BMI and ln(CRP)level was almost J-shaped for both genders at ages 40, 50 and 60 respectively. However, the ln(CRP) level

slightly increased at age 60 for both men and women. The predicted values of ln(CRP) level increased rapidly when BMI >30 and reach a maximum at BMI=50. The large confidence interval at the right tail of the BMI distributions is probably the effect of a small number of the individuals who are more obese (BMI>40).

Figure 5 presents a comparison of the FP2 model, the linear-quadratic model and the categorical model stratified by gender. The ln(CRP)-BMI curves were based on individuals who never smoked, aged 50 plus. The top panel of **Figure 5** evaluates FP2 model versus linear-quadratic model. Both models generate a J-shape. However, the linear-quadratic model underestimated the predicted values of ln(CRP) when BMI was between 20 to 35 for male and 20 to 38 for female respectively. It also overestimated the findings of ln(CRP) measure when BMI was over 34 and 38 for both men and women compared to the FP2 model.

The middle panel of **Figure 5** evaluates the FP2 model versus the linear model. Both models produce similar predicted values for ln(CRP) when BMI <35. However, linear regression
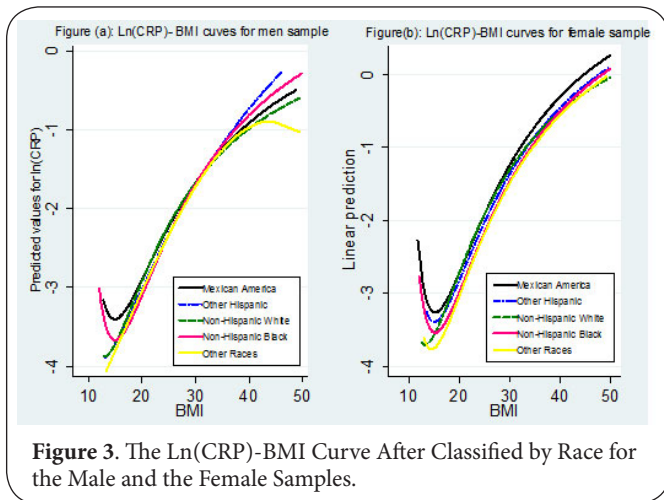

Figure 3. The Ln(CRP)-BMI Curve After Classified by Race for the Male and the Female Samples.

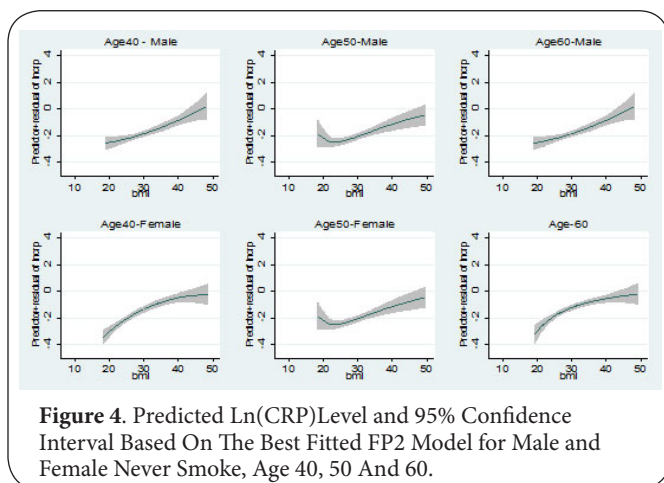**Figure 4**. Predicted Ln(CRP)Level and 95% Confidence Interval Based On The Best Fitted FP2 Model for Male and Female Never Smoke, Age 40, 50 And 60.
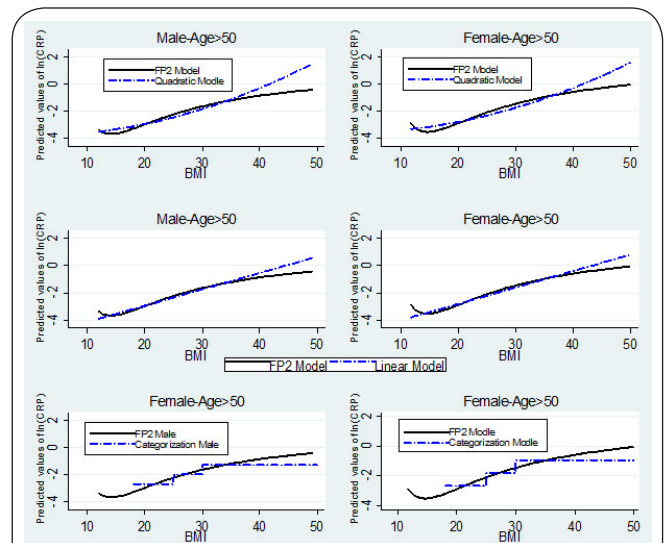

Figure 5. The Best-Fitted FP2 Model Compared With the Linear-Quadratic Model for BMI (Top Row); In the Middle row Regression Linear Model Compared With the Best-Fitted FP2 Model; and In The Bottom Row the Categorical Model Compared With the Best-Fitted FP2 Model for Participants who Never Smoke at Age>50, for Male and Female Samples.

model overestimated the predicted values of ln(CRP) when BMI was greater than 35 for the male and 36 for the female samples, respectively.

In the bottom panel of **Figure 5**, the FP2 model was compared with the categorical model for the respondents aged 50 plus, who never smoked. BMI was again categorized into the three groups recommended by the WHO. The categorical approach underestimated the findings of the ln(CRP) measure

when BMI was approximately more than 35 for both genders. It also overestimated the results when BMI was almost less than 20 for men and women, respectively. All analyses were repeated for all age participants and we obtained the same findings for male and female samples.

Ln(CRP) level estimates for male and female samples (unadjusted model) are presented in **Table 2**. For both samples, the FP2 model produced the lowest root mean square error (Root MSE) compared to the quadratic, linear and categorical model. The categorical approach yielded the highest root MSE. All models produced a positive association between CRP level and BMI measure. Nevertheless, the impact of BMI on CRP level was different based on the selected model. The categorical approach provided the highest association for both samples. The effect of an increase in BMI on elevated CRP level was slightly high on the female sample over the male sample for all models. In the linear model, the effect is constant which means that one unit increase in BMI will increase ln(CRP) level by β_3.

In the polynomial model, one unit increase at BMI will raise ln(CRP) level by $\beta_3(BMI^2)$. In the categorical model, one unit increase in BMI will elevate ln(CRP) level by β_3 if 25<BMI <30, and by if BMI >30. In the FP2 model, one unit increase in BMI increase ln(CRP) level by $\beta_4 X_{BMI}^{-2} + \beta_4 X_{BMI}^{-2} \log(\overline{BMI})$, where $(\overline{BMI}) =$ BMI/10. For example, at the male sample in the unadjusted models, if BMI increase from 45 to 46, ln(CRP) estimate level will increase by 0.009, 0.01, 0.11 and 0.22 for FP2, the quadratic, linear and categorical models respectively. The categorical and the linear models overestimated the results, while the quadratic model underestimates the findings compared to FP2 model. In addition, the Linear model produced a constant effect, for example, if BMI increase from 40 to 41, the Ln(CRP) level expected to increase by 0.11, and if BMI increase from 24 to 25, the elevated level of ln(BMI) expected to increase by 0.11. The FP2 model produced the most accurate findings, neither underestimating nor overestimating the elevated

level of CRP. The findings were similar for the male/female sub-samples with the female sample reporting a slightly higher effect compared to the male sample.

For the adjusted model, all models were adjusted for age, cotinine level, and alcohol consumption. The findings for each model were presented in **Table 3**. The same findings were obtained for both the male and the female sub-sample. FP2 model remains the best-fitted model over other models.

Discussion

In this study, the comparison between different statistical methods to measure the association between elevated CRP level and obesity was undertaken. All models found a positive association between elevated levels of CRP and Obesity and this relationship varied across gender and race. The findings in this study with regard to the association between the ln(CRP)-BMI association and estimate curves are matched with the findings of previous studies [15-21]. However, The FP method provided a better model fit compared to other statistical models (linear, quadratic, and categorical). The FP approach has also been shown to provide a more robust and precise method to determine the function of the BMI covariate [22]. In the FP models, the findings of the BMI-ln(CRP) curves were different across gender and race. In particular, BMI-ln(CRP) curves were a J-shaped for both genders (male and female) [7]; however, the BMI-ln(CRP) curves were slightly higher for female compared to male; and in the female sub-sample the curve for Mexican American was highest compared to all other race groupings. In the male sub-sample, the other-Hispanic group displayed the highest curve when BMI>30. We also found differences in the shape of BMI-ln(CRP) curve and ln(CRP) estimates compared to other models for women and men sub-samples.

Using the WHO categorization for obesity, the categorical model provided the least precise predicted values of ln(CRP) level [7,23]. Making full use of the scale data for BMI, the linear

Table 2. The Comparison Between FP2, Quadratic, Linear and Categorical Unadjusted Models.

	FP2		Quadratic	Linear	Categorical		
	BMI(P_1)	BMI(P_2)	BMI(2)	BMI	BMI<25	25<BMI<30	BMI>30
Male	-2.68	-17.98	0.197	0.11	---	0.75	1.24
Root MSE	1.225	--	1.251	1.233	---	1.261	--
Female	-2.76	-21.069	0.199	0.12	---	0.94	1.63
Root MSE	1.223	--	1.265	1.239	--	1.277	--

For the male sample: The BMI for FP2 model are calculated as follows

$$BMI(P1) = \left(\frac{BMI}{10}\right)^{-2} - 0.1508230081, and\ BMI(P_2) = \left(\frac{BMI}{10}\right)^{-2} \times \ln\left(\frac{BMI}{10}\right) + -0.1426520405.$$ For the Quadratic model the BMI(2) is measured as $\left(\frac{BMI}{10}\right)^2 - 6.630288127$. For the male sample: The BMI for FP2 model are calculated as: $BMI(P1) = \left(\frac{BMI}{10}\right)^{-2} - 0.1436105802$, and $BMI(P_2) = \left(\frac{BMI}{10}\right)^{-2} \times \ln\left(\frac{BMI}{10}\right) + -0.1393489324$. For the Quadratic model the BMI(2) is measured as $\left(\frac{BMI}{10}\right)^2 - 6.630288127$.

Table 3. The comparison between FP2, Quadratic, Linear and Categorical adjusted models.

	FP2		Quadratic	Linear	Categorical		
	BMI(P_1)	BMI(P_2)	BMI(2)	BMI	BMI<25	25<BMI<30	BMI>30
Male	3.166	-26.36	0.137	0.085	---	0.47	1.07
Root MSE	1.128	--	1.133	1.13	---	1.1539	--
Female	-1.662	-20.23	0.152	0.0987	---	0.70	1.48
Root MSE	1.136	--	1.157	1.144	--	1.167	--

For the male sample: The BMI for FP2 model are calculated as follows
$$BMI(P1) = \left(\frac{BMI}{10}\right)^{-2} - 0.1272052723 \quad and \quad BMI(P_2) = \left(\frac{BMI}{10}\right)^{-2} - 0.131145657. \text{ For the Quadratic}$$
model the BMI(2) is measured as $\left(\frac{BMI}{10}\right)^{2}$. For the female sample: The BMI for FP2 model are calculated
as follows $BMI(P1) = \left(\frac{BMI}{10}\right)^{-2} -.1223678208 and \quad BMI(P_2) = \left(\frac{BMI}{10}\right)^{-2} -.1285304995$. For the
Quadratic model the BMI(2) is measured as $\left(\frac{BMI}{10}\right)^{2}$. The model is adjusted for age, cotinine level, and
alcohol consumption.

model demonstrated advantages over the categorical model. However, to use the linear models, the normality assumption must be achieved, and if not the linear model may overestimate or underestimate the findings. Linear models also produce a constant impact, of the association between ln(CRP) and BMI across the scale; this might be imprecise as one unit increase in BMI from 20 to 21 might not have the same influence if BMI increase from 35 to 36 [7]. The quadratic model yield a non-linear association between BMI and ln(CRP); however, this relation was restricted to a certain transformation, which might not be the best choice of non-linear approach.

The FP method has a number of advantages over the other models [24]; FP is a flexible model that allows the data itself to determine the function form for BMI across 44 available function[8]; and the closed test is used to select the best fitted model that produced the best predicted values for elevated ln(CRP) level based on BMI measure. The estimates of ln(CRP)-BMI curves from the FP method shows symmetry over other models, this is an indicator of less variation on the predicted values of ln(CRP). The findings of the FP approach showed that ln(CRP)level increased exponentially for extremely obese participants and also we obtained the same results for individuals aged 40, 50 and 60 for both genders.

A limitation of this study is the exclusion of missing data rather than statistically addressing these data. The focus in this study is only on the association between the elevated level of CRP and obesity for both genders controlling for age, cotinine level, and alcohol consumption. However, some other important explanatory variables including socio-economic position might produce a better model fit. Finally, the findings of this study are based on only five cross-sectional observational studies pooled from NHANES dataset, as such the findings may using other trials (e.g., randomized clinical trials).

Conclusion

Assessing the association between the obesity and CRP level is vital, as obesity and CRP level (as an inflammation factor) have been linked to cardiovascular disease (CVD) either in childhood or in adulthood. We found a positive association between CRP level and obesity for all selected statistical models (linear, non-linear, categorical and FP model). Categorizing the continuous variable and using categorical models, is the least preferable method to examine the relationship between obesity and CRP levels. While providing better estimates than the categorical models, the linear model requires data linearity assumptions and it also assumes a constant influence of the covariate on the outcome. Polynomials models (e.g., quadratic and cubic models) are slightly flexible as they are nonlinear models and also there is no need to categories the continuous variable; however the powers of the polynomial models are limited. FP model produced the best estimates for ln(CRP) level-BMI relationship with the minimum deviation over the other models and as such may be useful in examining other health outcomes.

Authors' contributions

Authors' contributions	GAZ	KM
Research concept and design	✓	--
Collection and/or assembly of data	✓	--
Data analysis and interpretation	✓	--
Writing the article	✓	✓
Critical revision of the article	--	✓
Final approval of article	✓	✓
Statistical analysis	✓	--

References

1. Finkelstein EA, Khavjou OA, Thompson H, Trogdon JG, Pan L, Sherry B and Dietz W. **Obesity and severe obesity forecasts through 2030**. *Am J Prev Med*. 2012; **42**:563-70. | Article | PubMed

2. Fish JS, Ettner S, Ang A and Brown AF. **Association of perceived neighborhood safety with [corrected] body mass index**. *Am J Public Health*. 2010; **100**:2296-303. | Article | PubMed Abstract | PubMed FullText

3. Royston P, Altman DG and Sauerbrei W. **Dichotomizing continuous predictors in multiple regression: a bad idea**. *Stat Med*. 2006; **25**:127-41. | Article | PubMed

4. Altman DG, Lausen B, Sauerbrei W and Schumacher M. **Dangers of using "optimal" cutpoints in the evaluation of prognostic factors**. *J Natl Cancer Inst*. 1994; **86**:829-35. | PubMed

5. Kahan BC, Rushton H, Morris TP and Daniel RM. **A comparison of methods to adjust for continuous covariates in the analysis of randomised trials**. *BMC Med Res Methodol*. 2016; **16**:42. | Article | PubMed Abstract | PubMed FullText

6. Dupont WD. **Review of Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modeling Continuous Variables, by Royston and Sauerbrei**. *The Stata Journal*. 2010. **10**:297-302.

7. Wong ES, Wang BC, Garrison LP, Alfonso-Cristancho R, Flum DR, Arterburn DE and Sullivan SD. **Examining the BMI-mortality relationship using fractional polynomials**. *BMC Med Res Methodol*. 2011; **11**:175. | Article | PubMed Abstract | PubMed FullText

8. Sauerbrei W, Royston P. **Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials**. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1999; **162**:71-94. | Article

9. Sauerbrei W and Royston P. **Multivariable Fractional Polynomial Models, in International Encyclopedia of Statistical Science**. *Springer*. 2011; 899-902.

10. Sauerbrei W, Meier-Hirmer C and Benner A et al. **Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs**. *Computational Statistics & Data Analysis*. 2006; **50**:3464-3485. | Article

11. CDC. **National Health and Nutrition Examination Survey (NHANES) stored biologic specimens: Guidelines for proposals to use samples and proposed cost schedule**. *Fed Regist 74*. 2009; **138**:35872–5. | Website

12. National Center for Health Statistics: **National Health and Nutrition Examination Survey: Plan and Operations**. 1999–2010. | Pdf

13. National Center for Health Statistics: **National Health and Nutrition Examination Survey 1999–2010 questionnaires, datasets, and related documentation**. Hyattsville, MD: Centers for Disease Control and Prevention. 2012. | Website

14. National Center for Health Statistics: **The National Health and Nutrition Examination Survey 2005–2006. Survey operations manuals, brochures, consent documents**. 2009. | Website

15. Altman DG and Royston P. **The cost of dichotomising continuous variables**. *BMJ*. 2006; **332**:1080. | Article | PubMed Abstract | PubMed FullText

16. Visser M, Bouter LM, McQuillan GM, Wener MH and Harris TB. **Elevated C-reactive protein levels in overweight and obese adults**. *JAMA*. 1999; **282**:2131-5. | Article | PubMed

17. Hak AE, Stehouwer CD, Bots ML, Polderman KH, Schalkwijk CG, Westendorp IC, Hofman A and Witteman JC. **Associations of C-reactive protein with measures of obesity, insulin resistance, and subclinical atherosclerosis in healthy, middle-aged women**. *Arterioscler Thromb Vasc Biol*. 1999; **19**:1986-91. | Article | PubMed

18. Choi J, Joseph L and Pilote L. **Obesity and C-reactive protein in various populations: a systematic review and meta-analysis**. *Obes Rev*. 2013; **14**:232-44. | Article | PubMed

19. Aronson D, Bartha P, Zinder O, Kerner A, Markiewicz W, Avizohar O, Brook GJ and Levy Y. **Obesity is the major determinant of elevated C-reactive protein in subjects with the metabolic syndrome**. *Int J Obes Relat Metab Disord*. 2004; **28**:674-9. | Article | PubMed

20. Heilbronn LK, Noakes M and Clifton PM. **Energy restriction and weight loss on very-low-fat diets reduce C-reactive protein concentrations in obese, healthy women**. *Arterioscler Thromb Vasc Biol*. 2001; **21**:968-70. | Article | PubMed

21. Le-Ha C, Beilin LJ, Burrows S, Oddy WH, Hands B and Mori TA. **Gender and the active smoking and high-sensitivity C-reactive protein relation in late adolescence**. *J Lipid Res*. 2014; **55**:758-64. | Article | PubMed Abstract | PubMed FullText

22. Sauerbrei W, Royston P and Binder H. **Selection of important variables and determination of functional form for continuous predictors in multivariable model building**. *Stat Med*. 2007; **26**:5512-28. | Article | PubMed

23. Naggara O, Raymond J, Guilbert F, Roy D, Weill A and Altman DG. **Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms**. *AJNR Am J Neuroradiol*. 2011; **32**:437-40. | Article | PubMed

24. Royston P, Ambler G and Sauerbrei W. **The use of fractional polynomials to model continuous risk variables in epidemiology**. *Int J Epidemiol*. 1999; **28**:964-74. | PubMed