Original Article

Dietary phosphorus intake and serum prostate-specific antigen in non-prostate cancer American adults: A secondary analysis of the National Health and Nutrition Examination Survey (NHANES), 2003-2010

Guohua Zhu BS[1,2†], Chi Chen MD[3†], Bin Hu BD BS[1,4], Dongbo Yuan MS[1,5], Weimin Chen MS[3], Wei Wang MS[1], JiaMing Su MS[1], Zhangcheng Liu BS[1], Ke Jiao BS[1], Xiaoyue Chen BS[3], Weihong Chen BS[1], Wei Zhang MS[1], Fa Sun MD[1], Jianguo Zhu MD, PhD[1,2,4,5]

[1]*Department of Urology, Guizhou Provincial People's Hospital, Guiyang, Guizhou Province, China.*
[2]*Guizhou Medical University, Guiyang, Guizhou Province, China*
[3]*Guizhou University School of Medicine, Guiyang, Guizhou Province, China*
[4]*Zunyi Medical University, Zunyi, Guizhou Province, China*
[5]*Department of Immunology and Microbiology, Guiyang College of Traditional Chinese Medicine, Guiyang, Guizhou Province, China*
[†]*Both authors contributed equally to this manuscript*

Background and Objectives: Previous study has reported phosphorus intake is associated prostate cancer (PCa), but the association between phosphorus intake and serum prostate specific antigen (PSA) levels hasn't been reported in non-history of PCa population. Therefore, we performed a secondary data analysis based on existing data from the public Nutrition Examination Survey (NHANES) (2003-2010) database. **Methods and Study Design:** Totally 6403 participants were selected from NHANES (2003-2010) database. The interested independent and dependent variables were considered as dietary phosphorus intake and PSA level, respectively. Covariates included demographic data, dietary data, physical examination data, and comorbidities. Weighted linear regression and generalized additive models were used to addressing the linear and non-linear link of phosphorus intake to PSA level. **Results:** Linear association between phosphorus intake and PSA was not detected [β=0.016 (95% Confidence Interval (CI) -0.012, 0.045)]. But we found an existing nonlinearity. By the recursive algorithm, the inflection point was 1151 mg. On the left side of the inflection point, we did not find the correlation between dietary phosphorus intake (per 100 change) and PSA level [β=-0.04 (95% CI -0.11, 0.02), *p*=0.2155], while dietary phosphorus intake (per 100 change) positively associated with PSA [β=0.05 (95% CI 0.01, 0.09) *p*=0.0293] on the right side of inflection point. **Conclusions:** There is a non-linear correlation between dietary phosphorus intake and PSA. Dietary phosphorus intake was positively associated with increased PSA when dietary phosphorus intake is beyond 1151 mg after adjusting other covariates. Over 1151 mg per day dietary phosphorus intake may be the risk factor for PSA increasing.

Key Words: dietary phosphorus intake, prostate-specific antigen (PSA), National Health and Nutrition Examination Survey (NHANES)

INTRODUCTION

Prostate cancer (PCa), which has been regarded as the fifth most common cause of cancer-related death worldwide, about 1.3 million new cases and over 359000 deaths were appeared in the high-income countries,[1] it is still the most commonly diagnosed cancer in men all over the world. Dairy foods have generally been associated with an increased risk of PCa, a present mate-analysis showed that a summary relative risk for the highest increased risk of PCa compared with the lowest total dairy intakes.[2] Growing evidence have been suggested higher phosphorus intake is associated with PCa.[3-6] Phosphorus is a mineral found in dietary foods, it is independently associated with risk of lethal and high-grade PCa.[7,8] It has reported that there might be a positive association be-

Corresponding Author: Dr Jianguo Zhu, Department of Urology, Guizhou Provincial People's Hospital, The Affiliated Hospital of Guizhou University, Guiyang, Guizhou Province, 550002, China.
Tel: +86-0851-85925503; Fax: +86-0851-85925503
Email: doctorzhujianguo@163.com
Manuscript received 21 December 2019. Initial review completed 03 February 2020. Revision accepted 28 May 2020.
doi: 10.6133/apjcn.202007_29(2).0015

tween phosphorus intakes and risk of PCa.[9-11]

Prostate-specific antigen (PSA) has been widely used for screening tool for the disease as it is a useful tumor marker for PCa.[12,13] It is widely used in clinical practice as a screening tool for PCa since 1988 with a 24% positive predictive value.[14,15] However, current literature have indicated that many factors can affect the level of serum PSA, and lead to the uncertain effect on the quality of PCa screening. If phosphorus intake has a greater impact on PSA levels in the non-PCa US males, the detected PSA level may be overemphasized to induce possible overdiagnosis and treatment. Therefore, the focus of our study is the relationship between phosphorus intake and PSA level in non-PCa Americans.

In the present work, we performed a secondary data analysis based on existing data from the public Nutrition Examination Survey (NHANES) database (https://www.cdc.gov/nchs/nhanes).

METHODS

Data source

We compiled data from NHANES dataset of 2003-2010 (including five circles). NHANES is an ongoing cross-sectional observational study that collects health-related information from nationally representative samples of the civilian, noninstitutionalized population of the United States. The NHANES is conducted by the National Center for Health Statistics. All procedures for data collection were approved by the Center's ethics review board. All participants provided written informed consent before data collection. The detailed introduction can be download from CDC official website (link: https://wwwn.cdc.gov/nchs/nhanes/tutorials/default.aspx).

Participants selection

A total of 42470 participants were initially involved in NHANES of 2003 to 2010. After a series of screenings, a total of 6403 males were selected for final data analysis. We listed the screening procedure as follows: (1) Female (n=21685); (2) Male aged ≥40 years old (n=8454); (3) Male tumor patients (n=7776); (4) Male who had recently or previously used statins, thiazide diuretics, NSAIDS, androgen (n=7399); (5) Male with prostatitis, or recent prostate manipulation (i.e., rectal examination within 1 week, and prostate biopsy, surgery, or cystoscopy within 1 month) (n=7196); (6) Missing PSA data (n=6622); (7) Lack of phosphorus intake (n=6403). In the end, 6403 subjects were included in the study (Flowchart was shown in Figure 1 in detail). All participants provided informed consents both before the interview and examination stages.

Variables

In our study, the dependent variables were dietary phosphorus (mg) intake, the quantification and determination methods are detailed in the NHANES official website. The targeted independent variable was PSA (ng/mL). The measurement of phosphorus intake and serum PSA have been described in NHANES official website (link: https://www.cdc.gov/nchs/nhanes).

According to the published studies with discovery the association between phosphorus intake and PCa incidence, we selected the following covariates, which was associated with PSA level,[16,17] to establish multivariate models: continuous variables consisted of LDL-cholesterol (mg/dL), poverty income ratio (PIR), body mass index (kg/m^2), alcohol first day (mg), Vitamin D (ng/mL), C-reactive protein(mg/dL), Glycohemoglobin (%), HDL-cholesterol (mg/dL), age (year), protein intake first day (mg) and triglycerides (mg/dL). Furthermore, and the categorical variables as following in Table 1: hypertension history, diabetes history, coronary heart disease, stroke, education level, marital status, physical activity, and enlarged prostate. In short, covariates can be summarized as demographic data, dietary data, physical examination data and comorbidities in the NHANES database. A more detailed description of these variables can be found on the NHANES official website.

Statistical analysis

The statistical analysis was in guidelines of the CDC (https://wwwn.cdc.gov/nchs/nhanes/tutorials/default.aspx). The entire process of data analysis consists of four aspects. Because the distribution of PSA in the population is skewed, we used the log2 function to transform it, furthermore, due to the large amount of phosphorus in the diet, we used Per 100 change to evaluate the change in the effect value.

Missing data address

It was noted that most of covariates involved in this study have record miss (Supplemental table 1). Therefore, the appropriate handle of missing data can enhance the statistical power and reduced the bias caused by missing record. In the present study, the multiple multivariate imputations, and the sensitivity analysis of data distribution among pre and post-imputation were performed (Supplemental table 2).

Variation trend among different dietary phosphorus intake groups

The dietary phosphorus intake was separated into 4 groups according to different dietary phosphorus intake groups (Quartile). Weighted chi-square test (categorical variables), one–way ANOVA (normal distribution), or Kruskal-Wallis test (skewed distribution) were used to calculate the differences among four quartile groups.

Linear relationship between dietary phosphorus intake and PSA

The weighted univariate and multivariate linear regression model was used to exploreobvirous the linear relationship and constructed four models: Model I, no covariates were adjusted; Model II, only adjusted for demographics data (Race/Ethnicity); Poverty income ratio; Age (year); marital status, education level); Model III, Fully-adjusted for the covariates exhibited in Table 1; Model IV, a weighted generalized additive model (GAM).

Nonlinearity addressing

(1)We conducted the GAM model and smooth curve fitting (penalized spline method) to explore the nonlinearity association between dietary phosphorus intake and PSA levels. (2) If the GAM model detects nonlinearity, we

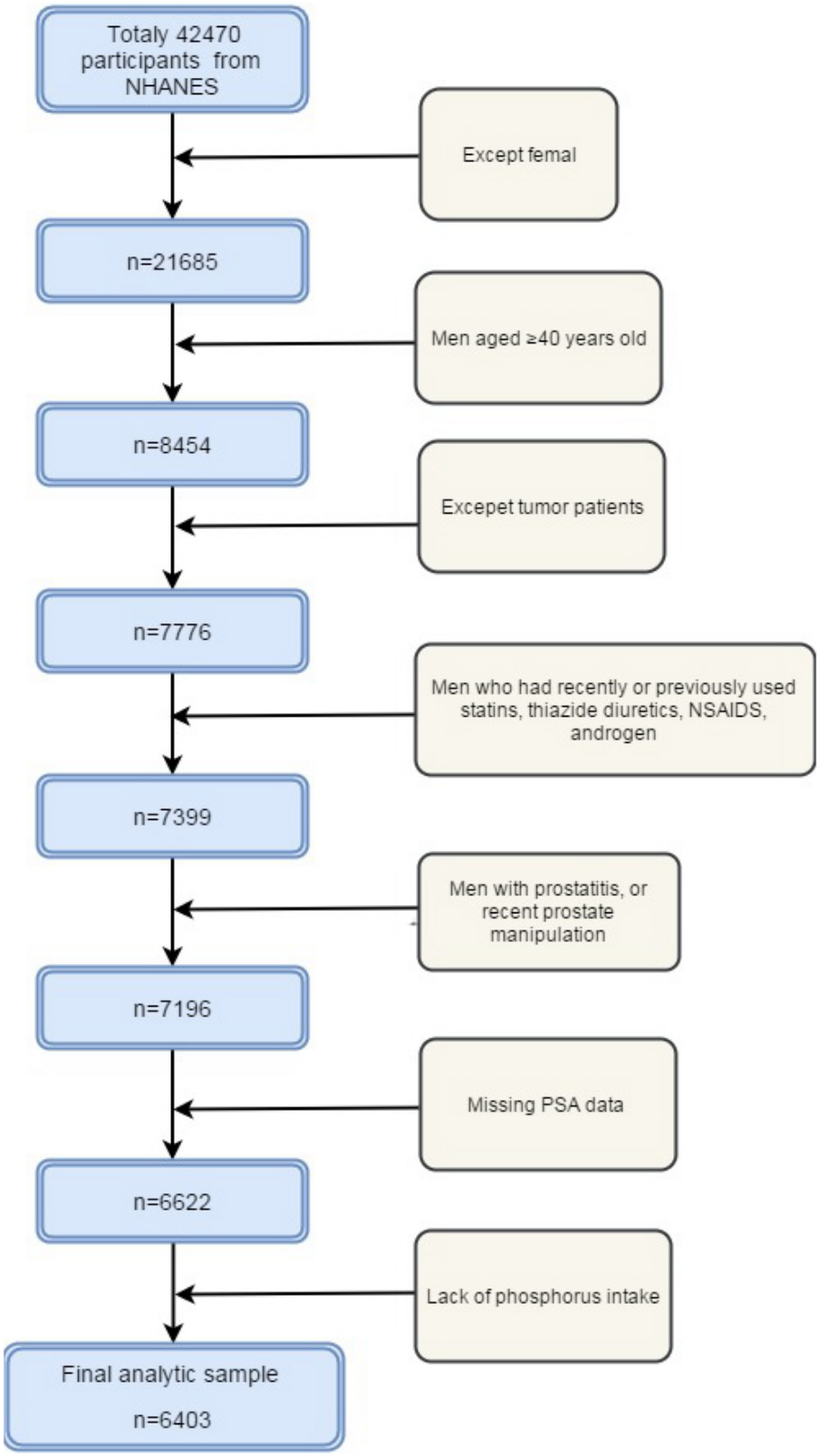


Figure 1. The flowchart of participants selection.

we first calculate the inflection point using a recursive algorithm, and then the piecewise linear fitting model was used to calculate the correlation between dietary phosphorus intake and PSA on both sides of the inflection point. (3) We determined the best fit model based on the p-value of the log-likelihood ratio test (linear regression model and two piecewise linear regression models).

Sensitivity analysis

(1) Change the continuous variable dietary phosphorus intake into the classified variable (quartile), and observe p for trend, its purpose is to observe whether the change of the effect value is isometric, and the possibility of the existence of the curve relationship can be preliminarily determined. (2) Because linear regression model can't handle nonlinear relationships, therefore, we use the

Table 1. Baseline characteristics of selected participants

Phosphorus (mg)	Q1 (57.0-962)	Q2 (963-1327)	Q3 (1328-1778)	Q4 (1779-6152)	*p*-value
n	1334	1135	1135	1138	
Laboratory data					
PSA (ng/mL) log2 transform	0.12 (-3.84-5.32)	0.14 (-3.84-5.23)	-0.05 (-3.84-5.32)	-0.29 (-3.84-5.32)	<0.001**
Vitamin D (mcg)	55.6 (21.9)	60.1 (21.8)	62.2 (21.4)	64.0 (21.5) 62.3	<0.001**
LDL-cholesterol (mg/dL)	116 (37.1)	118 (36.5)	120 (34.1)	119 (34.0)	0.263
Alcohol (gm) first day	8.71 (23.8)	13.2 (31.4)	14.4 (32.7)	20.9 (42.6)	<0.001**
C-reactive protein(mg/dL)	0.22 (0.01-18.5)	0.20 (0.01-17.5)	0.19 (0.01-13.7)	0.17 (0.01-10.50)	<0.001**
Glycohemoglobin (%)	5.99 (1.25)	5.94 (1.23)	5.88 (1.12)	5.78 (1.06)	<0.001**
HDL	48.9 (14.9)	49.4 (14.6)	47.6 (13.8)	48.3 (14.9)	0.018*
Triglycerides (mg/dL)	129 (30.0-1438)	129 (23.0-2566)	142 (21.0-1223)	139 (24.0-2693)	<0.001**
Demographics data					
Poverty income ratio	2.36 (1.53)	2.67 (1.60)	2.96 (1.62)	3.00 (1.65)	<0.001**
Body mass index, kg/m^2	28.5 (5.54)	28.4 (5.48)	29.1 (6.02)	29.3 (5.71)	<0.001**
Smoked at least 100 cigarettes in life					0.0005**
Yes	61.5	60.7	56.9	55.2	
No	38.5	39.3	43.1	44.9	
Age, year	63.0 (12.5)	61.3 (12.7)	58.0 (12.3)	54.1 (10.9)	<0.001**
Race/Ethnicity					<0.0001**
Mexican American	6.38	6.74	6.38	6.99	
Other Hispanic	4.62	3.51	3.02	2.62	
Non-Hispanic white	64.4	73.6	78.3	81.8	
Non-Hispanic black	17.1	11.1	8.58	6.07	
Other race – including multi-racial	7.54	5.06	3.67	2.48	
Education level					<0.0001**
Less than high school	13.7	8.91	5.56	5.28	
High school	40.8	38.8	34.1	34.6	
More than high school	45.5	52.3	60.3	60.1	
Marital status					<0.0001**
Married	66.1	69.2	76.1	71.6	
Single	30.3	26.2	18.9	21.8	
Living with partner	3.61	4.48	4.90	6.68	

Mean +/− SD for continuous variables: *p* value was calculated by weighted linear regression model.
% for categorical variables: *p* value was calculated by weighted chi-square test.
*p<0.05, **p<0.01.

Table 1. Baseline characteristics of selected participants (cont.)

Phosphorus (mg)	Q1 (57.0-962)	Q2 (963-1327)	Q3 (1328-1778)	Q4 (1779-6152)	*p*-value
Questionnaire data					
Hypertension history					0.0001**
yes	43.2	41.3	38.1	33.9	
no	56.8	58.7	62.0	66.1	
Diabetes history					0.0007**
yes	15.9	13.5	11.6	9.82	
no	84.1	86.5	88.4	90.2	
Coronary heart disease					<0.0001**
yes	10.1	8.27	5.00	3.67	
no	89.9	91.7	95.0	96.3	
Stroke					<0.0001**
yes	6.04	3.63	3.31	1.66	
no	94.0	96.4	96.7	98.3	
Physical activity					0.0009**
Sits	26.4	23.9	24.3	20.9	
Walks	50.7	47.9	49.1	44.7	
Light loads	15.1	19.2	16.3	24.0	
Heavy work	7.85	9.01	10.3	10.4	
Prostate conditions					0.1901
yes	15.2	17.2	13.7	14.7	
no	84.8	82.8	86.3	85.3	

Mean ± SD for continuous variables: p value was calculated by weighted linear regression model.
% for categorical variables: *p* value was calculated by weighted chi-square test.
*$p<0.05$, **$p<0.01$.

GAM model to adjust the continuous variables in covariates as curves and observe the changes of the effect values.

All analysis was performed using statistical software R (http://www.r-project.org, The R Foundation) and EmpowerStats (http://www.empower-stats.com, X&Y Solutions, Inc., Boston, MA). A *p* value of less than 0.05 (two-sided) was considered statistically significant.

RESULTS

Baseline characteristics of participants

The weighted distribution of selective participants socio demographic characteristics and other covariates for the selected NHANES 2003-2010 population is shown in Table 1. The average age of the participants was 51.1±9.6 years old. There was no statistically significant difference in smoked at least 100 cigarettes in life, prostate conditions and LDL-cholesterol (all *p* values >0.05). Comparing Q1 with Q2 group, subjects with high dietary phosphorus intake were younger, had higher Vitamin D intake, Poverty income ratio, Body mass index (kg/m^2), Alcohol (mg) first day, Protein (mg) first day, Triglycerides (mg/dL), Education level. The subjects had lower C-reactive protein (mg/dL), glycohemoglobin (%), lower incidence of hypertension, diabetes, coronary heart dis ease, stroke. Most of the participants were non-hispanic white population.

Univariate and multivariate analysis

The results of the univariate and multivariate linear regression model are showed in Table 2. In the non-adjusted model. It was showed that the PSA level is decreased by -0.020 (-0.025, -0.015) with per 100 change of dietary phosphorus intake increases. The association between phosphorus intake and PSA levels was no significant difference after adjusted for demographics variables (minimally-adjusted) (*p*=0.53938), adjusted for all covariates in table 1 (*p*=0.71416), and GAM to adjust continuous variables in covariates as non-linear (*p*=0.48592).

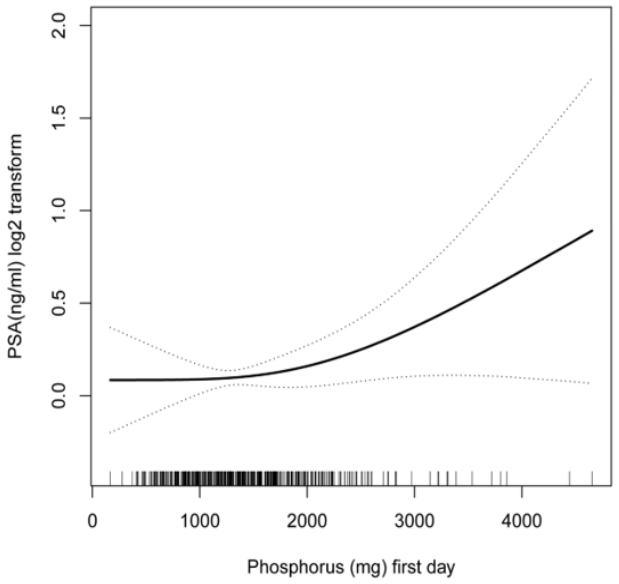


Figure 2. Results of standard linear regression model and two-piecewise linear regression model: there is a non-linear association between dietary phosphorus and PSA.

To explore whether there is a nonlinear relationship between dietary phosphorus intake and PSA. In a sensitivity analysis, dietary phosphorus intake was converted into categorical variable according to quartile and estimated phosphorus for trend (Table 2). We found that the phosphorus for trend was consistent with phosphorus intake as a continuous variable, and the trend of effect values among different phosphorus intake groups were Non-isometric variation. These findings indicated the possibility of nonlinearity on phosphorus intake and PSA level.

Identification of non-linear relationship

By generalized additive model and smooth curve fitting (penalized spline method), the non-linear association between dietary phosphorus and PSA. We first calculated inflection point for further explaining the nonlinearity by recurrence function (inflection point was 1151 mg). And estimated effect sizes trend on both sides of inflection points (p for the log-likelihood ratio test is 0.025) (Figure 2). Due to the limitations of methodology, we were unable to pool the nonlinear relationships among the post-imputation data. Pre-imputation data was used to explore the dietary phosphorus intake and PSA, the nonlinear trends of the data among pre- and post- imputation were approximately the same in sensitivity analysis (Supplemental Figure), therefore we only put the curve fitting done by the data pre-imputation. In the study.

In two-piecewise linear regression model and recursive algorithm we calculated the inflection point was 1151 mg, on the left of the inflection point, Effect values are defined as 95%CI and p values were -0.04 (log2 transformation) (-0.11, 0.02) and *p*=0.2155, and on the right side of the inflection point, 95%CI and *p* values were 0.05 (log2 transformation) (0.01, 0.09), *p*=0.0293 (Table 3). Which shows a positive association between dietary phosphorus intake and PSA. It means that there was somewhat U-shape between dietary phosphorus intake and PSA with dietary phosphorus intake threshold level of 1151 mg.

DISCUSSION

In the study, we reported the relationship between dietary phosphorus intake and PSA in American non-PCa population with age over 40 years old. Although we do not find the linear association between them ave threshold-effect. Our findings demonstrated that when phosphorus intake exceed 1151 mg, per 100 mg of phosphorus intake increases are independent associated with the 1.035 ng/mL increases of PSA level.

The previous study has reported that positive associations between phosphorus intake and lethal, advanced-stage, high-grade, and grade 7 PCa,[7] it is difficult to separate their effects on the high correlation between calcium and phosphorus intake caused.[18] In a trial cohort, total calcium and phosphorus intake was associated with an increased risk of total PCa.[19] Furthermore, Kapur S,[20] showed in their study that low dietary intake of phosphorus leads to an increase in serum concentration of 1,25-(OH)2-D for reduced risk of PCa in aging men. However, there is a case-control study conducted that there was no material association of dietary intake phosphorus with PCa risk.[21]

Table 2. Result of univariate and multivariate analysis by weighted linear regression model and GAM model

Exposure	Model I Non-adjusted model β (95% CI) *p*-value†	Model II Minimally-adjusted model β (95% CI) *p*-value†	Model III Fully-adjusted model β (95% CI) *p*-value†	Model IV GAM model β (95% CI) *p*-value†
Dietary phosphorus intake per 100 change	-0.020 (-0.025, -0.015) <0.00001	0.000 (-0.005, 0.006) 0.87117	0.016 (-0.012, 0.045) 0.25822	0.020 (-0.006, 0.046) 0.13734
Q1	0	0	0	0
Q2	-0.028 (-0.132, 0.076) 0.60065	0.064 (-0.038, 0.166) 0.21995	0.075 (-0.564, 0.714) 0.81768	-0.198 (-0.843, 0.447) 0.54943
Q3	-0.213 (-0.317, -0.109) 0.00006	-0.008 (-0.113, 0.096) 0.87400	-0.242 (-0.844, 0.359) 0.43237	-0.505 (-1.110, 0.100) 0.10772
Q4	-0.379 (-0.483, -0.275) <0.00001	-0.011 (-0.118, 0.096) 0.84144	0.212 (-0.382, 0.807) 0.48620	-0.093 (-0.676, 0.489) 0.75467
p for trend	<0.00001**	0.53938	0.71416	0.48592

95% CI: 95% confidence interval; Indicated effect sizes (β) were combined by Rubin's rule.
Model I: Adjust for none;
Model II: Adjusted for demographics data (race/ethnicity); poverty income ratio; age (year); marital status, education level);
Model III: Fully-adjusted for the covariates exhibited in Table 1
Model IV: All continuous variables in the covariates were adjusted as smooth
† *p*-value of GAM model
* $p<0.05$, ** $p<0.01$.

Table 3. Result of nonlinearity addressing by weighted two-piecewise linear model

	PSA (ng/mL) log2 transform
Fitting by weighted linear regression model	0.020 (-0.006, 0.046) 0.13734 †
Fitting by weighted two-piecewise linear regression model	
Inflection point	15.51
<15.51	-0.04 (-0.11, 0.02) 0.2155 †
≥15.51	0.05 (0.01, 0.09) 0.0293†
Log likelihood ratio test	0.025*

PSA: prostate-specific antigen; GAM: generalized additive model.
Independent variable is dietary phosphorus (mg) intake and dependent variable is PSA (ng/mL log2 transform).
Covariates involved in this model was the same as GAM model presented in Table 2
† *p*-value of GAM model
* $p<0.05$, ** $p<0.01$.

From 1992, the American Urological Association (AUA) and the American Cancer Society (ACS) recommended annual screening for men 50 years and older, PSA screening was widely adopted and was associated with increases in PCa incidence.[22] But PSA is synthesized in healthy prostate tissue, in benign prostatic hypertrophy (BPH) and in prostate cancer of all grades and stages.[23-25] Our study has showed that over 1151mg per day dietary phosphorus intake may be the risk factor for PSA increasing, and an article from based on NHANES shows that between 2001 and 2014, dietary phosphorus intake increased from 1345 to 1399 mg/day.[26] The previous clarified the relationship between dietary phosphorus intake and PCa, and PCa is currently screened mainly through PSA, but PSA can also be produced in healthy prostate tissue, and many factors that can affect the level of serum PSA.[27] These may affect the quality of prostate screening, and even lead to misdiagnosis or missed diagnosis. If there is a relationship between dietary phosphorus intake and PSA in non-Prostate cancer population, we conducted linear and nonlinear regression model to increase comparability, and the results revealed the possibility of a nonlinear relationship was detected. And used the GAM to elucidate the non-linear relationship, and finally calculated the inflection point by the recursive algorithm and discovered the saturation effect by two-piecewise linear regression.

There are some strengths of note: Firstly, we have a large sample size. Secondly, we used multiple interpolation to improve statistical efficiency and avoid the bias caused by the lack of data. Thirdly, this study clarifies the nonlinear relationship, which provides a new vision for the future study of the etiological mechanism of cancer, and provides some help for the formulation of corresponding public health policies. Finally, Sensitivity analysis ensured the robustness of the results.

Our study still has limitations: Firstly, due to the nature of cross-sectional study we used, that can provided only weak evidence between dietary phosphorus intake and PSA in normal non-tumor population, which was impossible to obtain the causal inference, however, when PSA is used to screen prostate cancer, the clinical scenario is also cross-sectional, so the results are still applicable. Secondly, our study is limited to the American, so it may not be suitable for Asia and other races. Furthermore, NHANES is a cross-sectional study, which means that it may did not include the variables that can affect the PSA level but not adjusted in our study. Finally, we can only adjust measurable confounding factors as in all observational studies, but there are a lot of other unfathomable confounding factors that we can't adjust.

Conclusion

There is a relationship between dietary phosphorus intake and PSA in normal non-tumor population, dietary phosphorus intake is positively correlated with PSA when phosphorus is over than 1151 mg. But it still needs large prospective clinical trials with robust methodology to confirm.

AUTHOR DISCLOSURES

The authors declare no conflict of interest.

This work was supported by grants from The National Natural Science Foundation of China (81660426), the Science and Technology Project of Guizhou Province in 2017([2017]5803), The High-level innovative talent project of Guizhou Province in 2018 ([2018]5639), The Science and Technology Plan Project of Guiyang in 2019 ([2019]2-15), Science and Technology Plan Project of Guizhou Province in 2019 ([2019]1203).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424. doi: 10.3322/caac.21492.
2. Gao X, LaValley MP, Tucker KL. Prospective studies of dairy product and calcium intakes and prostate cancer risk: a meta-analysis. J Natl Cancer Inst. 2005;97:1768-77. doi: 10.1093/jnci/dji402.
3. Allen NE, Key TJ, Appleby PN, Travis RC, Roddam AW, Tjønneland A et al. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. Br J Cancer. 2008;98:1574-81. doi: 10.1038/sj.bjc.6604331.
4. Rohrmann S, Platz EA, Kavanaugh CJ, Thuita L, Hoffman SC, Helzlsouer KJ. Meat and dairy consumption and subsequent risk of prostate cancer in a US cohort study. Cancer Causes Control. 2007;18:41-50. doi: 10.1007/s10552-006-0082-y.
5. Park Y, Mitrou PN, Kipnis V, Hollenbeck A, Schatzkin A, Leitzmann MF. Calcium, dairy foods, and risk of incident and fatal prostate cancer: the NIH-AARP Diet and Health Study. Am J Epidemiol. 2007;166:1270-9. doi: 10.1093/aje/kwm268.
6. Koh KA, Sesso HD, Paffenbarger RS Jr, Lee IM. Dairy products, calcium and prostate cancer risk. Br J Cancer. 2006;95:1582-5. doi: 10.1038/sj.bjc.6603475.
7. Wilson KM, Shui IM, Mucci LA, Giovannucci E. Calcium and phosphorus intake and prostate cancer risk: a 24-y follow-up study. Am J Clin Nutr. 2015;101:173-83. doi: 10.3945/ajcn.114.088716.
8. Chan JM, Pietinen P, Virtanen M, Malila N, Tangrea J, Albanes D et al. Diet and prostate cancer risk in a cohort of smokers, with a specific focus on calcium and phosphorus (Finland). Cancer Causes Control. 2000;11:859-67. doi: 10.1023/a:1008947201132.
9. Tavani A, Bertuccio P, Bosetti C, Talamini R, Negri E, Franceschi S, Montella M, La Vecchia C. Dietary intake of calcium, vitamin D, phosphorus and the risk of prostate cancer. Eur Urol. 2005;48:27-33. doi: 10.1016/j.eururo.2005.03.023.
10. Mitrou PN, Albanes D, Weinstein SJ, Pietinen P, Taylor PR, Virtamo J, et al. A prospective study of dietary calcium, dairy products and prostate cancer risk (Finland). Int J Cancer. 2007;120:2466-73. doi: 10. 1002/ijc.22553.
11. Tseng M, Breslow RA, Graubard BI, Ziegler RG. Dairy, calcium, and vitamin D intakes and prostate cancer risk in the National Health and Nutrition Examination Epidemiologic Follow-up Study cohort. Am J Clin Nutr. 2005;81:1147-54. doi: 10.1093/ajcn/81.5.1147.
12. Sirovich BE, Schwartz LM, Woloshin S. Screening men for prostate and colorectal cancer in the United States: does practice reflect the evidence?. JAMA. 2003;289:1414-20. doi: 10.1001/jama.289.11.1414.
13. Cooperberg MR, Moul JW, Carroll PR. The changing face of prostate cancer. J Clin Oncol. 2005;23:8146-51. doi: 10.1200/JCO.2005.02.9751.

14. Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360: 1310-9. doi: 10.1056/NEJMoa0810696.
15. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360: 1320-8. doi: 10.1056/NEJMoa0810084.
16. Wallis CJ, Satkunasivam R. Prostate cancer: Risk factors - you find what you are looking for. Nat Rev Urol. 2017;14: 202-4. doi: 10.1038/nrurol.2016.267.
17. Chu DI, Freedland SJ. Metabolic risk factors in prostate cancer. Cancer. 2011;117:2020-3. doi: 10.1002/cncr.25749.
18. Giovannucci E, Liu Y, Stampfer MJ, Willett WC. A prospective study of calcium intake and incident and fatal prostate cancer. Cancer Epidemiol Biomarkers Prev. 2006; 15:203-10. doi: 10.1158/1055-9965.EPI-05-0586.
19. Kesse E, Bertrais S, Astorg P, Jaouen A, Arnault N, Galan P, et al. Dairy products, calcium and phosphorus intake, and the risk of prostate cancer: results of the French prospective SU.VI.MAX (Supplémentation en Vitamines et Minéraux Antioxydants) study. Br J Nutr. 2006;95:539-45. doi: 10. 1079/bjn20051670.
20. Kapur S. A medical hypothesis: phosphorus balance and prostate cancer. Cancer Invest. 2000;18:664-9. doi: 10.3109/ 07357900009032832.
21. Tavani A, Bertuccio P, Bosetti C, Talamini R, Negri E, Franceschi S et al. Dietary intake of calcium, vitamin D, phosphorus and the risk of prostate cancer. Eur Urol. 2005; 48:27-33. doi: 10.1016/j.eururo.2005.03.023.
22. Potosky AL, Miller BA, Albertsen PC, Kramer BS. The role of increasing detection in the rising incidence of prostate cancer. JAMA. 1995;273:548-52.
23. Lilia H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat Rev Cancer. 2008;8:268-78. doi: 10.1038/nrc2351.
24. Herrala AM, Porvari KS, Kyllönen AP, Vihko PT. Comparison of human prostate specific glandular kallikrein 2 and prostate specific antigen gene expression in prostate with gene amplification and overexpression of prostate specific glandular kallikrein 2 in tumor tissue. Cancer. 2001; 92:2975-84. doi: 10.1002/1097-0142(20011215)92:12<297 5::aid-cncr10113>3.0.co;2-k.
25. Lintula S, Stenman J, Bjartell A, Nordling S, Stenman UH. Relative concentrations of hK2/PSA mRNA in benign and malignant prostatic tissue. Prostate. 2005;63:324-9. doi: 10. 1002/pros.20194.
26. McClure ST, Chang AR, Selvin E, Rebholz CM, Appel LJ. Dietary sources of phosphorus among adults in the United States: Results from NHANES 2001-2014. Nutrients. 2017; 9:95. doi: 10.3390/nu9020095.
27. Stephan C, Lein M, Jung K, Schnorr D, Loening SA. The influence of prostate volume on the ratio of free to total prostate specific antigen in serum of patients with prostate carcinoma and benign prostate hyperplasia. Cancer. 1997; 79:104-9.

Supplemental table 1. Nonlinearity addressing by weighted two-piecewise linear model

Variables	Non-missing	Missing
Vitamin D (μg)	6437	185
LDL-cholesterol (mg/dL)	3053	3569
Race/Ethnicity	6622	0
Poverty income ratio	6187	435
Body mass index, kg/m^2	6509	113
Alcohol (g) first day	6344	278
HDL	6622	0
C-reactive protein(mg/dL)	6621	1
Hypertension history	3939	2683
Diabetes history	3838	2784
Coronary heart disease	3864	2758
Stroke	3864	2758
Triglycerides (mg/dL)	6603	19
Enlarged prostate	4788	1834
Age (year)	6622	0
Lead (umol/L)	6617	5
Cadmium (nmol/L)	6617	5
Mercury, total (umol/L)	6617	5
Smoked at least 100 cigarettes in life	3854	2768

PSA: prostate-specific antigen; GAM: generalized additive model.
Independent variable is dietary phosphorus (mg) intake and dependent variable is PSA (ng/mL log2 transform).
Covariates involved in this model was the same as GAM model presented in Table 2.
$^{*}p<0.05$, $^{**}p<0.01$.

Supplemental table 2. Univariate and multivariate analysis by weighted linear regression model and GAM model

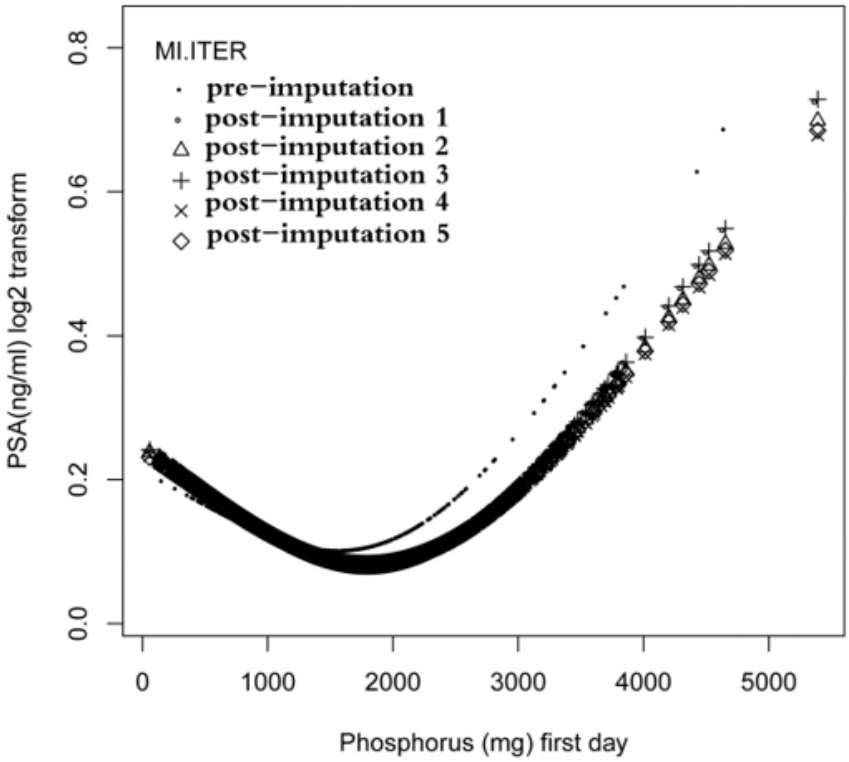
MI.ITER	0	1	2	3	4	5	*p*-value	*p*-value*
Laboratory data								
Vitamin D (mg)	60.21 (21.72)	60.23 (21.79)	60.27 (21.69)	60.28 (21.66)	60.29 (21.70)	60.28 (21.71)	1.000	1.000
LDL-cholesterol (mg/dL)	118.60 (35.22)	119.78 (35.67)	119.46 (35.67)	118.81 (35.57)	119.16 (35.90)	119.32 (35.59)	0.581	0.270
Alcohol (g) first day	14.40 (34.34)	14.31 (34.45)	14.23 (34.37)	14.48 (34.49)	14.42 (34.28)	14.43 (34.27)	0.999	0.968
C-reactive protein(mg/dL)	0.19 (0.01-18.50)	0.19(0.01-18.50)	0.19 (0.01-18.50)	0.19 (0.01-18.50)	0.19 (0.01-18.50)	0.19 (0.01-18.50)	1.000	1.000
Glycohemoglobin (%)	5.89 (1.17)	5.89 (1.17)	5.89 (1.17)	5.89 (1.17)	5.89 (1.17)	5.89 (1.17)	1.000	1.000
HDL	48.58 (14.37)	48.58 (14.37)	48.58 (14.37)	48.58 (14.37)	48.58 (14.37)	48.58 (14.37)	1.000	1.000
Triglycerides (mg/dL)	133.00 (21.00-2693.00)	133.00 (-295.03-2693.00)	133.00 (-75.64-2693.00)	133.00 (-175.92-2693.00)	133.00 (-208.16-2693.00)	133.00 (-93.90-2693.00)	1.000	1.000
Demographics data								
Poverty income ratio	2.74 (1.61)	2.72 (1.61)	2.73 (1.62)	2.73 (1.63)	2.72 (1.63)	2.73 (1.62)	0.976	0.995
Body mass index, kg/m^2	28.75 (5.55)	28.77 (5.55)	28.76 (5.55)	28.76 (5.55)	28.75 (5.56)	28.76 (5.55)	1.000	1.000
Age, year	59.47 (12.79)	59.47 (12.79)	59.47 (12.79)	59.47 (12.79)	59.47 (12.79)	59.47 (12.79)	1.000	1.000
Race/Ethnicity, n (%)							1.000	-
Mexican American	1239 (18.71)	1239 (18.71)	1239 (18.71)	1239 (18.71)	1239 (18.71)	1239 (18.71)		
Other Hispanic	387 (5.84)	387 (5.84)	387 (5.84)	387 (5.84)	387 (5.84)	387 (5.84)		
Non-Hispanic White	3504 (52.91)	3504 (52.91)	3504 (52.91)	3504 (52.91)	3504 (52.91)	3504 (52.91)		
Non-Hispanic Black	1241 (18.74)	1241 (18.74)	1241 (18.74)	1241 (18.74)	1241 (18.74)	1241 (18.74)		
Other Race - Including Multi-Racial	251 (3.79)	251 (3.79)	251 (3.79)	251 (3.79)	251 (3.79)	251 (3.79)		
Education level, n (%)							1.000	-
Less than high school	1155 (17.47)	1157 (17.47)	1158 (17.49)	1159 (17.50)	1160 (17.52)	1157 (17.47)		
High school	2488 (37.63)	2493 (37.65)	2493 (37.65)	2491 (37.62)	2490 (37.60)	2494 (37.66)		
More than high school	2969 (44.90)	2972 (44.88)	2971 (44.87)	2972 (44.88)	2972 (44.88)	2971 (44.87)		
Marital status, n (%)							1.000	-
Married	4493 (67.94)	4500 (67.96)	4500 (67.96)	4498 (67.93)	4500 (67.96)	4500 (67.96)		
Single	1788 (27.04)	1790 (27.03)	1790 (27.03)	1791 (27.05)	1790 (27.03)	1790 (27.03)		
Living with partner	332 (5.02)	332 (5.01)	332 (5.01)	333 (5.03)	332 (5.01)	332 (5.01)		

The results show that there is no significant difference in the data before and after interpolation, that is, interpolation does not cause the change of data distribution.

Supplemental table 2. Univariate and multivariate analysis by weighted linear regression model and GAM model (cont.)

MI.ITER	0	1	2	3	4	5	*p*-value	*p*-value*
Questionnaire data								
Hypertension history, n (%)							0.221	-
yes	1671 (42.4)	3076 (46.5)	3080 (46.5)	3064 (46.3)	3198 (48.3)	3202 (48.4)		
no	2268 (57.6)	3546 (53.6)	3542 (53.5)	3558 (53.7)	3424 (51.7)	3420 (51.7)		
Diabetes history, n (%)							0.191	-
yes	626 (16.3)	1351 (20.4)	1298 (19.6)	1319 (19.9)	1318 (19.9)	1387 (21.0)		
no	3212 (83.7)	5271 (79.6)	5324 (80.4)	5303 (80.1)	5304 (80.1)	5235 (79.1)		
Coronary heart disease, n (%)							0.063	-
yes	283 (7.32)	833 (12.6)	795 (12.0)	795 (12.0)	803 (12.1)	770 (11.6)		
no	3581 (92.7)	5789 (87.4)	5827 (88.0)	5827 (88.0)	5819 (87.9)	5852 (88.4)		
Stroke							0.082	-
yes	180 (4.66)	425 (6.42)	482 (7.28)	482 (7.28)	469 (7.08)	515 (7.78)		
no	3684 (95.3)	6197 (93.6)	6140 (92.7)	6140 (92.7)	6153 (92.9)	6107 (92.2)		
Smoked at least 100 cigarettes in life, n (%)							0.609	-
yes	2431 (63.1)	4108 (62.0)	4104 (62.0)	4055 (61.2)	4092 (61.8)	4100 (61.9)		
no	1423 (36.9)	2514 (38.0)	2518 (38.0)	2567 (38.8)	2530 (38.2)	2522 (38.1)		
Physical activity, n (%)							0.077	-
Sits	991 (26.2)	1508 (22.8)	1548 (23.4)	1607 (24.3)	1588 (24.0)	1629 (24.6)		
Walks	1867 (49.3)	3227 (48.7)	3183 (48.1)	3195 (48.3)	3243 (49.0)	3239 (48.9)		
Light loads	617 (16.3)	1253 (18.9)	1263 (19.1)	1220 (18.4)	1189 (18.0)	1161 (17.5)		
Heavy work	309 (8.17)	634 (9.57)	628 (9.48)	600 (9.06)	602 (9.09)	593 (8.95)		
Prostate conditions, n (%)							0.809	-
yes	836 (17.6)	1180 (17.8)	1142 (17.3)	1125 (17.0)	1127 (17.0)	1145 (17.3)		
no	3928 (82.5)	5442 (82.2)	5480 (82.8)	5497 (83.0)	5495 (83.0)	5477 (82.7)		

The results show that there is no significant difference in the data before and after interpolation, that is, interpolation does not cause the change of data distribution.



Supplemental figure 1. The correlation between dietary phosphorus and PSA. Different line patterns indicated different data sources (pre- or post-imputation).