Neural network approach to predict the association between blood cadmium levels and hypertension

Kisok Kim1*, Hyejin Park2
1College of Pharmacy, Keimyung University, Daegu 42601, Republic of Korea; kimkisok@kmu.ac.kr (K.K.).
2Department of Health Sciences, Dongduk Women’s University, Seoul 02748, Republic of Korea; hjpark@dongduk.ac.kr (H.P.).

Abstract: Hypertension is a global health concern and a major risk factor for cardiovascular disease. Early prevention and management based on risk prediction is a principal goal of many national health policies. We studied the relationship between blood cadmium concentrations and hypertension and developed an artificial neural network (ANN) that predicts hypertension risk. For this study, we utilized data from the Korean National Health and Nutrition Examination Survey (KNHANES), conducted between 2008 and 2013, which is a nationwide population-based survey of the Korean population. We extracted and analyzed sociodemographic characteristics, serum cadmium levels, and blood pressure information from a sample of adults aged 19 years and above (n=11,530). After adjusting for sociodemographic factors, cadmium levels were positively associated with the risk of hypertension (p < 0.001). Groups with high cadmium levels significantly increased the odds ratios for hypertension compared to the lowest tertile. An ANN model in which sociodemographic factors and the blood concentration of cadmium were the principal inputs yielded a predictive accuracy of 0.773 and an area under the curve of 0.823. ANNs with appropriate inputs can identify population subgroups at high risk of developing hypertension and will aid in the formulation of policies that prevent disease.

Keywords: Artificial neural network, Blood concentration, Cadmium, Hypertension, KNHANES, Risk prediction.

1. Introduction

Hypertension is a major public health problem worldwide and one of the main risk factors for cardiovascular disease. Hypertension’s global clinical and economic burdens are high and continue to increase; in 2015, hypertension caused 17.7 million deaths worldwide [1, 2]. In Korea, the mortality rate from hypertension in 2020 was 2.0%. It was the ninth leading cause of death, and its prevalence in those aged 30 years and older was 34.2% [3]. The Korean National Health and Nutrition Examination Survey (KNHANES), a nationwide population-based survey conducted by the Korean Centers for Disease Control and Prevention, usefully identifies factors that cause hypertension given the big data on blood pressure [4-6]. Well-known risk factors for hypertension include a genetic predisposition and certain lifestyle factors, but heavy metal exposure may also play an important role [7-9].

Previous studies have indicated a potential association between cadmium (Cd) exposure and hypertension. Upon exposure to Cd, its concentration in the blood serves as a biomarker for the level of exposure. Cd is primarily stored in soft tissues, with the liver and kidneys being the major sites of accumulation [10]. A population-based cohort study by Gambelunghe, et al. [11] revealed an association between lead exposure and blood pressure, as well as hypertension. Moreover, a meta-analysis of eight studies conducted by Navas-Acien, et al. [9] and Martins, et al. [12] suggested a link between arsenic exposure and hypertension. The 2011-2018 US National Health and Nutrition Examination Survey found an association between mercury exposure and the prevalence of hypertension among Asian populations in the US [13]. Furthermore, a study in Korea’s general population revealed a correlation between high
blood manganese levels and hypertension [14]. Exposure to Cd has been reported to increase hypertension in several epidemiological and experimental studies [12, 13]. A prospective study conducted in the US reported a positive association between low to moderate levels of urinary Cd and hypertension in the general population [14]. Various mechanisms have been reported to elucidate the link between Cd exposure and blood pressure, with tubular epithelial cell damage recognized as a prominent consequence of Cd exposure [15]. The impairment of renal function caused by tubular dysfunction may influence the relationship between Cd levels and blood pressure [12].

Healthcare research has widely applied artificial neural network (ANN) models recently. ANNs can derive highly accurate nonlinear correlations between independent and dependent (outcome) variables [16]. Recently, they have become increasingly used for disease classification given the massive computational resources available, the big data of large surveys, and the precise measurements [17, 18]. Such data efficiently trains machine-learning models that detect trends and patterns when screening and classifying many diseases [19, 20]. We thus sought a correlation between the body burden of Cd and hypertension and developed an ANN for predicting hypertension in Korean adults; we used KNHANES data to these ends.

2. Methods

2.1. Study Population

Data were obtained from the 2008–2013 KNHANES, a nationwide survey conducted by the Korean Centers for Disease Control and Prevention. The KNHANES Website (https://knhanes.kdca.go.kr/knhanes/) contains the survey data that is available to the general public. The KNHANES utilizes a stratified, multistage cluster-sampling procedure; the proportional allocation is based on National Census Registry information. In total, 11,530 adults aged at least 19 years were included in the analysis, after excluding those for whom adequate blood pressure and blood Cd data were lacking and who did not answer the relevant questionnaire items. The protocol was approved by the Korean Ministry of Health and Welfare and the research adhered to the principles of the Declaration of Helsinki. All participants provided written, informed consent.

2.2. Data Collection

We extracted data on demographics (age, sex), socioeconomic status (income, educational level), and lifestyle habits (physical exercise, smoking status, and alcohol consumption) from the KNHANES database. We obtained anthropometric measurements such as height and weight, from participants wearing lightweight clothing without shoes. We calculated the Body Mass Index (BMI) by dividing the weight (kg) by the square of the height (m). Based on the World Health Organization (WHO) guidelines for Asian populations, participants were categorized as underweight (BMI < 18.5 kg/m²), normal weight (18.5 ≤ BMI < 23.0 kg/m²), overweight (23.0 ≤ BMI < 25.0 kg/m²), or obese (BMI ≥ 25.0 kg/m²). Cd levels were measured using Zeeman effect graphite furnace atomic absorption spectrophotometry (Perkin-Elmer AAnalyst 600, Turku, Finland) with a limit of detection (LOD) of approximately 0.30 μg/L. For participants with blood Cd concentrations below the LOD, a detection limit value divided by the square root of 2 was assigned. We measured blood pressure on the right arm three times using a mercury sphygmomanometer and recorded the average of the second and third measurements. Hypertension was defined as a systolic blood pressure (SBP) ≥ 140 mmHg and/or a diastolic blood pressure (DBP) ≥ 90 mmHg, or the taking of medication to treat hypertension [21].

2.3. ANN Model Development

The data were pre-processed using the StandardScaler function of the Python scikit-learn library. The data were split into training and test sets at a ratio of approximately 7:3. The ANN included input layer, 12 hidden layers, and 1 output layer. We employed a stochastic gradient descent (SGD) optimizer with the ADAM variant for training the ANN. The learning rate was set to 0.001, and the training process spanned 100 epochs. Rectified linear unit (ReLU) activation functions were used in all hidden layers, while
the output layer utilized a sigmoid function. To prevent overfitting, dropout regularization (30%) was applied to the input layer. Categorical cross-entropy served as the loss function, reflecting the binary classification nature of the problem. The model was implemented using TensorFlow v1.12.0. Finally, we evaluated the model’s predictive performance using receiver operating characteristic (ROC) curve analysis.

2.4. Statistical Analysis

We used the Mantel-Haenszel chi-square test to compare between-group differences in categorical variables and the two-sample t test to compare differences in continuous variables. We calculated the geometric means with 95% confidence intervals (CIs) of blood concentration of Cd using the antilog values of the mean, natural-log-transformed values. We employed geometric means on a normal probability plot to enhance the resemblance to a normal distribution. We derived odds ratios (ORs) for hypertension, with 95% CIs (compared to the reference values), via multivariate logistic regression analyses that incorporated all covariates. The Cochrane-Armitage test for trends was used to explore the linearity of the relationship between the Cd level and hypertension prevalence. The statistical analyses considered the unique features of the survey design; appropriate procedures employed weighted data. SAS version 9.4 software (SAS Institute Inc., Cary, NC, USA) performed all analyses.

3. Results

The mean age of the 11,530 participants was 44.9 years (standard deviation: 15.0 years); 49.2% were male and 26.3% were hypertensive. Age, sex, household income, education level, body mass index, and alcohol consumption differed significantly between the participants with and without hypertension (all $p < 0.001$). The geometric mean blood level of Cd was significantly higher in the hypertension group ($p < 0.001$) (Table 1).

Table 1. The sociodemographic characteristics and blood level of cadmium by hypertension (HTN) status in Korean adults aged ≥ 19 years.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>HTN</th>
<th>Non-HTN</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years (SD)</td>
<td>44.9 (15.0)</td>
<td>55.5 (12.7)</td>
<td>41.1 (13.9)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Male</td>
<td>5,676 (49.2)</td>
<td>1,730 (15.0)</td>
<td>3,946 (34.2)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5,854 (50.8)</td>
<td>1,303 (11.3)</td>
<td>4,551 (39.5)</td>
<td></td>
</tr>
<tr>
<td>Income, $US/month (SD)</td>
<td>3,052 (2.183)</td>
<td>2,532 (2.108)</td>
<td>3,237 (2.180)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>≤ Elementary school</td>
<td>2,091 (18.1)</td>
<td>1,061 (9.2)</td>
<td>1,030 (8.9)</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>1,215 (10.6)</td>
<td>453 (4.0)</td>
<td>762 (6.6)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>4,393 (38.1)</td>
<td>927 (8.0)</td>
<td>3,466 (30.1)</td>
<td></td>
</tr>
<tr>
<td>≥ College</td>
<td>3,831 (33.2)</td>
<td>592 (5.1)</td>
<td>3,239 (28.1)</td>
<td></td>
</tr>
<tr>
<td>BMI, kg/m² (SD)</td>
<td>23.7 (3.4)</td>
<td>25.1 (3.3)</td>
<td>23.2 (3.2)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Regular exercise, n (%)</td>
<td>2.638 (22.9)</td>
<td>660 (21.8)</td>
<td>1,978 (23.3)</td>
<td>0.088</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td>2.900 (25.2)</td>
<td>746 (24.6)</td>
<td>2,154 (25.4)</td>
<td>0.411</td>
</tr>
<tr>
<td>Alcohol consumer, n (%)</td>
<td>6.718 (58.3)</td>
<td>1,995 (65.8)</td>
<td>4,723 (55.6)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Cadmium (µg/L), geometric mean (95% CI)</td>
<td>0.94 (0.93-0.95)</td>
<td>1.15 (1.13-1.17)</td>
<td>0.88 (0.87-0.89)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Note: *p*-values were calculated using the t-test or the Mantel-Haenszel chi-squared test.

Table 2 lists the prevalences of hypertension by the tertiles of blood concentration of Cd. The prevalences and ORs correlated positively with increasing Cd levels ($p$ for trend < 0.001). Compared to
subjects in the lowest tertile of Cd (<0.76 μg/L), the crude hypertension OR was 3.09 (95% CI, 2.71–3.51) among those with Cd > 1.25 μg/L. This association remained unchanged after adjustment for age and sex (OR = 1.64; 95% CI, 1.41–1.90; \( p \) for trend < 0.001) (model 1), and indeed after adjustment for all potential risk factors (OR = 1.82; 95% CI, 1.17–2.83; \( p \) for trend < 0.001) (model 2).

Table 2.
Prevalences and odds ratios (with 95% CIs) of hypertension when the two highest tertiles were compared to the population with the lowest tertile of blood level of cadmium among Korean adults ≥ 19 years of age.

<table>
<thead>
<tr>
<th>Risk measurement</th>
<th>Blood level of cadmium</th>
<th>( p ) for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tertile 1 (&lt; 0.76 μg/L)</td>
<td>Tertile 2 (0.76–1.25 μg/L)</td>
</tr>
<tr>
<td>Prevalence %</td>
<td>5.33</td>
<td>8.96</td>
</tr>
<tr>
<td>Odds ratio ( a )</td>
<td>Crude (Reference)</td>
<td>2.04 (1.77–2.34)</td>
</tr>
<tr>
<td>Model 1</td>
<td>1.00 (Reference)</td>
<td>1.32 (1.13–1.53)</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.00 (Reference)</td>
<td>1.32 (1.13–1.55)</td>
</tr>
</tbody>
</table>

Note: \( a \) Model 1: Adjusted for age and sex. Model 2: Additionally adjusted for income, educational status, BMI, regular exercise, and cigarette smoking and alcohol consumption status.

Based on the logistic regression data, we developed an ANN with nine inputs, including sociodemographic characteristics and blood levels of Cd (Figure 1). After normalizing all variables, the ANN model achieved an accuracy of 0.773. The area under the ROC curve (AUC) was 0.823 (Figure 2).

Figure 1.
Schematic diagram of ANN model.
4. Discussion

In this nationwide cross-sectional study that enrolled adult Koreans aged at least 19 years, a positive association was apparent between the blood level of Cd and hypertension. After taking into account all the other factors, the odds ratios for high blood pressure went up linearly as Cd levels went up. This is in line with what other studies have found, which is that high Cd levels are linked to high blood pressure [13, 22]. In addition, as smoking is intricately linked to blood Cd levels and poses a potential risk factor for cardiovascular disease [23, 24], we performed subgroup analyses focusing on smoking. The results showed that current smoking may make the link between high blood pressure and Cd levels stronger, with a fully adjusted OR of 1.97 (95% CI, 1.42–2.74) among people in the highest Cd exposure tertile. Although the underlying mechanism remains poorly understood, disruption of calcium homeostasis, increased oxidative stress, and vascular endothelium impairment may all be involved [25–27].

This study utilized blood Cd concentrations as a biomarker to assess the body burden of Cd. Epidemiologic research commonly uses blood or urine Cd concentrations as primary biomarkers to evaluate exposure and internal dose. However, it is important to consider that Cd's half-life differs between blood and urine [28]. Urinary and blood Cd levels increase proportionately to the stored Cd in the body, reflecting cumulative Cd exposure in both biomarkers. Moreover, blood Cd, with a half-life of 3–4 months, also provides insight into recent exposure [29, 30]. Furthermore, considering the absorption route of Cd is crucial, as the bloodstream absorbs inhaled Cd more efficiently than ingested Cd [29]. Therefore, blood and urinary Cd levels may provide distinct information regarding the timing and source of Cd exposure [31]. Nevertheless, blood cadmium levels correlated well with urinary cadmium levels [32, 33].
Associations between blood Cd levels and hypertension in the general population were reported in several studies [34, 35]. On the other hand, some studies have found no association between blood Cd levels, blood pressure, or hypertension [36, 37]. Disparities in the geographic, ethnic, and socioeconomic backgrounds of the participants involved in each study may account for these differences in research findings. Furthermore, it implies that using whole blood as a biomarker for assessing exposure or body burden could affect the strength of the association. Cross-sectional studies utilizing blood as a potential biomarker for Cd exposure may have limited applicability in terms of temporal interpretation. Furthermore, there were inconsistencies in the definition of hypertension across the studies. The present study classified participants as hypertensive if they were using antihypertensive medications and had elevated systolic and/or diastolic blood pressure. However, some studies did not consider the use of antihypertensive drugs as one of the criteria for defining hypertension, and others only considered "history of hypertension" for the classification of hypertension.

ANNs represent a recent and notable advancement in the field of nature-inspired algorithms. ANNs emulate the structure and function of the human brain and offer a powerful tool for analyzing intricate and non-obvious relationships among diverse predictors. By employing suitable neural network architectures and optimizing training weights, ANNs facilitate the prediction of medical outcomes. The medical field primarily developed ANNs and widely uses them for disease diagnosis, prognosis, and clinical decision-making [38]. Recently, researchers have used ANNs to predict the development of various diseases, such as diabetes and heart failure [39, 40]. Recently, they have become recognized as powerful tools for predicting the impacts of multiple variables with complex inter-relationships on specific outcomes [41]. ANNs identify complex, nonlinear relationships between dependent and independent variables and all interactions among predictors [42].

Several attempts have been made to use ANNs to predict hypertension, employing various datasets. Most recent studies have reported AUCs from 0.64 to 0.82, lower than our 0.823, with the exception of one that reported a remarkable AUC of 0.96 [43-47]. One possible explanation of the among-study disparities is that most prior ANNs employed only sociodemographic factors and/or comorbidities as inputs; we added the blood level of Cd. As Cd status significantly affects hypertension, that inclusion may have enhanced the AUC. Our ANN predicts hypertension based on both Cd levels and sociodemographic parameters. Our study is subject to several limitations. First, the cross-sectional design of the data precluded the establishment of a causal relationship between blood Cd levels and hypertension. Second, the single measurement of blood Cd may not accurately reflect long-term exposure to Cd. Third, despite the incorporation of demographic and lifestyle factors into the ANN model, unmeasured confounders such as the severity of hypertension, family history, complications, and renal function could still have influenced the outcomes. Fourth, the ANN model has a tendency to overfit, making it difficult to optimize parameters for optimal accuracy. Consequently, we selected parameters that yielded the best results among the models we evaluated. ANNs can serve as an effective adjunctive tool for clinical decision-making, enabling the identification of adults at high risk for hypertension and facilitating early intervention. To enhance prediction performance and improve precision and accuracy, future studies should incorporate a broader range of input data, including detailed population variables and pathological factors, and apply the proposed model trained on datasets from more diverse populations.

5. Conclusions

In this population-based study of 11,530 participants, the blood level of Cd significantly predicted the hypertension risk. We developed an ANN as a predictive model to estimate hypertension risk; the inputs included both the Cd levels and the sociodemographic characteristics of the KNHANES databases. Our ANN could be useful for identifying individuals at risk of developing hypertension, allowing for early treatment and possible prevention. However, we must validate our model in other clinical settings. However, our model must be validated in other clinical settings. Also, further studies employing different ANN architectures and input features are required to improve predictive accuracy and precision.
Funding:
This research is supported by the Ministry of Education, Korea (Grant number: NRF-2022R1F1A10744681220982085990102, NRF-2022R1A2C1006963).

Institutional Review Board Statement:

Transparency:
The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Competing Interests:
The authors declare that they have no competing interests.

Authors’ Contributions:
Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

Copyright:
© 2024 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

References

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 4: 336-344, 2024
DOI: 10.55216/25768484/v8i4.978
© 2024 by the authors; licensee Learning Gate


S. K. Park, Z. Zhao, and B. Mukherjee, "Construction of environmental risk score beyond standard linear models using machine learning methods: Application to metal mixtures, oxidative stress and cardiovascular disease in NHANES," *Environmental Health*, vol. 10, pp. 1-17, 2017. [https://doi.org/10.1016/j.s12940-017-0310-9](https://doi.org/10.1016/j.s12940-017-0310-9)

X. Gao, X. Xin, Z. Li, and W. Zhang, "Predicting postoperative pain following root canal treatment by using artificial neural network evaluation," *Scientific Reports*, vol. 11, no. 1, p. 17243, 2021. [https://doi.org/10.1038/s41598-021-96777-8](https://doi.org/10.1038/s41598-021-96777-8)


