Published in final edited form as:

J Occup Environ Med. 2009 December ; 51(12): 1422–1436. doi:10.1097/JOM.0b013e3181bf6c8d.

Bone Lead Level Prediction Models and Their Application to Examining the Relationship of Lead Exposure and Hypertension in the Third National Health and Nutrition Examination Survey (NHANES-III)

Sung Kyun Park, Sc.D.^{1,*}, Bhramar Mukherjee, Ph.D.^{2,*}, Xi Xia², David Sparrow, D.Sc.³, Marc G. Weisskopf, Ph.D., Sc.D.⁴, Huiling Nie, Ph.D.⁴, and Howard Hu, M.D., Sc.D.^{1,4}

¹Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA

²Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA

³VA Normative Aging Study, Veterans Affairs Boston Healthcare System and the Department of Medicine, Boston University School of Medicine, Boston, MA, USA

⁴Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

Abstract

Objective—We developed prediction models for bone lead using blood lead levels and other standard covariates in a community-based cohort of older men.

Methods—Participants having bone lead levels measured by K-x-ray fluorescence were included in the model selection process (n=825). Predictors of each tibia and patella lead were identified in three quarters of the population and then predicted the bone lead levels in the remaining one quarter and in the Community Lead Study.

Results—18 predictors were selected for tibia (blood lead, age, education, occupation, smoking status, pack-years of cigarette, serum levels of phosphorus, uric acid, calcium, creatinine and total and high-density lipoprotein cholesterol, hematocrit, body mass index, systolic and diastolic blood pressure, diagnoses of cancer and diabetes; $R^2=0.32$) and 16 for patella lead (among the predictors included in the tibia model diagnosis of cancer, serum levels of calcium and total cholesterol were not included in patella lead model, but diagnosis of hypertension was included; $R^2=0.34$), respectively. The correlation coefficients between the observed and predicted values were 0.43–0.50 for tibia and 0.52–0.58 for patella lead in internal and external validation. We applied these predicted bone lead models to the Third National Health and Nutrition Examination Survey (NHANES-III) to examine associations with hypertension and found relatively more significant associations compared to blood lead.

Conclusions—This study suggests that the prediction equations may be used to predict bone lead levels in other community-based cohorts with reasonable accuracy.

Address Correspondence to: Sung Kyun Park, SPH II-M6240, Department of Environmental Health Sciences, University of Michigan School of Public Health, 109 S. Observatory Street, Ann Arbor, MI 48109, Phone: 734-936-1719, Fax: 734-763-8095, sungkyun@umich.edu.

*These authors contributed equally to this work.

Introduction

Environmental exposure to lead in the United States has decreased substantially since the 1970s with an 84 percent decline in geometric mean blood levels of lead in adults (1,2). However, there still remain serious concerns regarding the effects of cumulative exposure to lead, particularly, in association with chronic health endpoints (3). Lead levels in bone have been suggested as an indicator of cumulative lead exposure (3,4). The half-life of lead in bone is of the order of years to decades (5,6). In adults, approximately 95% of the total body burden of lead is deposited in the skeleton (7,8) and therefore, bone lead levels may better predict the effects of lead toxicity that arise from long-term low to moderate exposure. The evolution of K x-ray fluorescence (KXRF) technology has made in vivo measurements of bone lead concentrations possible (3,4). Numerous studies have reported associations of elevated bone lead levels as measured by KXRF with various chronic health outcomes (9-11). The KXRF measurements require fairly sophisticated trained personnel and technology and thus, accurate bone lead measurements are not commonly available in standard community-based cohorts due to limitations in infrastructure and budgetary resources. On the other hand, lead in blood reflects, for the most part, recent lead exposure and is relatively easier and less expensive to measure and therefore, may be more frequently available in other cohorts.

Several studies have reported determinants of bone lead levels (12-19). These studies presented several multivariate regression models for bone lead levels in terms of covariates such as age, race, smoking and drinking behavior, and education. However, all these studies, except the study by Kosnett et al. (19), did not contain a model for bone lead level in terms of blood lead and other covariates, and did not address the issue of predictability of bone lead levels based on a fitted equation.

The purpose of this study is to develop and validate a prediction equation for chronic lead exposure as measured in tibia (representing cortical bone) and patella (representing trabecular bone) by KXRF using blood lead levels and other demographic, socio-economic, and clinical variables that are commonly available in any study cohort of lead. Additionally, we illustrate how predicted bone lead measures can add to prediction of health outcomes in a new cohort after incorporating the uncertainty in the predictions.

Materials and Methods

Study subjects

In this study we focused on participants of the Normative Aging Study (NAS) established by the Veterans Administration in 1961. The NAS is a longitudinal study in which 2280 community-dwelling men from 21 to 80 were recruited from the Greater Boston area. The enrollment contained both veterans and nonveterans, and represented a great variety of occupations, while only few, if any participants were employed in primary lead industry. Based on current status, physical examination and medical history, qualified volunteers were selected. Those excluded consisted of ones with history or presence of chronic conditions such as heart disease, hypertension, diabetes mellitus, cancer, peptic ulcer, gout, recurrent asthma, bronchitis, or sinusitis. Criteria of disqualification also included systolic blood pressure (SBP) >140 mmHg or diastolic blood pressure (DBP) >90 mmHg, but not biochemical parameters. During the follow-up period, study subjects returned every 3-5 years, took extensive examinations and had laboratory, anthropometric and questionnaire data collected.

Starting in 1991, assistants at Department of Veterans Affairs Outpatient clinic in Boston took fresh whole blood specimen from participants during their regular tests. Permission was sought to use these subjects for KXRF bone lead measurements.

Blood lead measurement

The blood samples for lead measurement were analyzed by graphite furnace atomic absorption with Zeeman background correction (ESA laboratories, Inc., Chelmsford, MA). Values below the minimum detection limit of 1 µg/dL (<1%) were coded as 0. The instrument was calibrated with National Institute of Standards and Technology Standard Reference Material (NIST SRM 955a, lead in blood) after every 20 samples. Ten percent of samples were run in duplicate; at least 10% of the samples were controls and 10% were blanks. In tests on reference samples from the Centers for Disease Control and Prevention (Atlanta, GA), the coefficient of variation (CV) ranged from 8% for concentrations from 10 to 30 µg/dL to 1% for higher concentrations. In this study, the CV was 5% for concentrations below 10 µg/dL. Compared with a NIST target of 5.7 µg/dL, 24 measurements by this method gave a mean of 5.3 (SD=1.2) µg/dL.

KXRF bone lead measurements

Bone lead measurements were taken of each subject's mid-tibial shaft and patella using a KXRF instrument (ABIOMED Inc., Danvers, MA). The tibia and patella have been targeted for bone lead research because they consist mainly of cortical and trabecular bone, respectively. The physical principles and technical details of this instrument have been described elsewhere (20,21). The unit of measurement derived is µg of lead per g of bone mineral (µg/g). The instrument provides a continuous unbiased point estimate that oscillates around the true bone lead value, thus, negative measurements are sometimes produced when bone-lead values are close to zero. The instrument also provides an estimate of the uncertainty associated with each measurement that is equivalent to a single standard deviation. Retaining all the measurements, including the negative ones, as opposed to applying a threshold approach makes better use of data in epidemiologic studies (22).

Statistical analysis

The initial dataset contained 1786 observations on 877 NAS participants. To control the quality of KXRF measurement, we first identified tibia and patella lead measurements with estimated measurement uncertainty greater than 10 and 15 µg/g, and discarded these observations. Additionally, the KXRF instrument used for bone lead measurement had been updated in 1999. To ensure homogeneity among observations, the observations measured by the new instrument were excluded from our working dataset. We initially included 28 potential predictors in our analysis, a list of which is provided in Table 1. We deleted observations with severe missing data in the initial set of covariates. After the data-cleaning procedure, our working dataset then consisted of 825 subjects with a total of 1,434 observations, with subjects having more than one observation in most cases. Among the study subjects only one had four observations; 126 had three observations; 354 had two observations and 344 had only one observation.

For this study, each individual's first measurements (baseline measurements, n=825) were included in the development of prediction models, instead of fitting a longitudinal model to the entire set, as (i) prediction is a more difficult concept in a longitudinal set-up, and (ii) because our goal is to develop a prediction equation which is usable in other cohorts, which may not have longitudinal data available. The follow-up observations were used for validating our models fitted on the baseline measurements. Figure 1 demonstrates procedures of model development and validation.

Analysis with baseline dataset and development of prediction model—The 825 subjects were randomly assigned to either a training set or a test set with an approximate ratio of 3:1 after deleting the missing observations in covariates or response. The training set was used for model selection and fitting the prediction equation, whereas the test set was used to assess the prediction performance of the fitted model on a set of independent observations, not used in the model building process.

Among the 28 potential predictors, we pre-selected three variables considered to be important in determining lead burden in the body from previous studies: age, job type and education (13,18,23,24). To identify other important predictors, we first calculated the residuals corresponding to a linear regression model of bone lead on the three pre-selected variables, fitted on the training set. We treated these residuals as the response in the next model selection procedure step and performed the least absolute shrinkage and selection operator (LASSO), a novel model selection plus shrinkage estimation procedure (25), instead of traditional stepwise methods. In usual least squares linear regression models, the regression coefficients are estimated by minimizing the residual sum of squares. However, if a large number of predictors are available, and some variables are highly correlated, then variance of the ordinary least squares estimators may be unacceptably high. The LASSO operator, proposed by Tibshirani (25), estimates a vector of regression coefficients by minimizing the residual sum of squares, subject to sum of the absolute value of the regression coefficients being less than a constant. Because of the nature of the constraint, the LASSO shrinks some of the regression coefficients to 0, or eliminates those variables. The LASSO has better prediction accuracy than subset selection when there may be a large number of predictors with moderate to small effects, which appears to be the case in this situation. The model fitted by LASSO was then assessed by several traditional model selection criteria, like the Akaike information criterion (AIC), the Bayes information criterion (BIC), Mallows' Cp and adjusted R^2 .

To minimize the sensitivity of the variable selection process with respect to a single random partition of the data to training set and test set, we repeated the selection procedures in 1000 randomly partitioned test and training sets and created a list of predictors which were selected in more than 50% of the replications. We then fitted a final linear regression model with these selected predictors on our fixed initial training set. We computed partial R^2 corresponding to each predictor using type-II sum of squares which evaluates the partial effect of a predictor by comparing a model with the full set of covariates to a model of dimension one less, by removing that particular predictor, but retaining all other covariates.

Finally, we verified the predictive ability of the fitted models on two test sets. Test set 1 consisted of approximately 25% of the baseline observations with complete case-data, not included in the training set. Test set 2 consisted of the next follow-up observation available on the subjects included in Test set 1. Thus, both the test sets are completely non-overlapping with the training set on which fitted equations are derived. We calculated predicted values for bone lead levels corresponding to the Test set observations based on the fitted equation from the training set. We then calculated the correlation between observed bone lead levels and predicted bone lead levels in the test sets. We also calculated descriptive statistics on the absolute prediction error ($|\text{observed}-\text{predicted}|$) distribution to assess prediction accuracy. In order to determine relative accuracy of the predicted bone lead levels, we compared the absolute prediction errors to the estimated measurement uncertainties which are generated by the KXRF instrument and represent inherent variability of the bone lead measurements.

We also constructed a set of reduced models without blood pressure and disease diagnoses. The reason for constructing this secondary model is the following. Predicted bone lead levels from our full model, if naively used in an association model corresponding to similar health outcomes as response, could potentially lead to strongly significant results that are artifact of the model building process and duplicate use of the outcome data in prediction and association. Hereafter, we refer to the model with all predictors including blood pressure and disease diagnoses as “full model” and the model without those variables as “reduced model”.

Other variables that have been reported to be associated with bone lead levels, such as dietary intakes of calcium and vitamin D (12), were also considered, but not included in the final model because their influences were minimal and because they might not be routinely collected in

population studies which precludes investigators from using this prediction model. We also considered cumulative traffic as a surrogate of past lead exposures to traffic particles from combustion of leaded gasoline (26), but did not include it because of the same reason above.

Development of prediction models with NTx—We also developed prediction models including urinary cross-linked N-telopeptides of type I collagen (NTx), because NTx is considered a specific marker of bone resorption (27) and thus may reflect lead mobilization from bone (4,28). An inverse of NTx ($1/\text{NTx}$) and an interaction between blood lead and $1/\text{NTx}$ were included in the models in order to account for the effect modification of blood lead by the rate of bone turnover. Due to missing in NTx responses, smaller numbers of subjects were used for the models with NTx.

External validation—To validate the prediction model in another cohort, we used data from the Community Lead Study (CLS) that includes minority and female subjects. Details of population characteristics can be found elsewhere (13). Briefly, subjects were recruited from the pool of subjects who participated in a study funded by the National Institutes of Health (NIH; “Impact of Sleep-Disordered Breathing in Older Adults”, NIH HL51075; PI, D. Sparrow). Between 1999 and 2000, 84 minority subjects aged ≥ 35 years (mean age 50 years, 82% African American, 67% female) were invited to measure bone and blood lead levels with the same techniques used in the NAS. Characteristics of the CLS population are presented in Table 2. Because all variables selected in the training set were not available in the CLS, we constructed a prediction equation using a subset of six predictors that were available in the CLS and had high partial R^2 in the prediction model: age, blood lead levels, education, job-type, cumulative cigarette pack-years and smoking status. Hereafter, we refer to this model as “6-predictor model”. We carried out similar validation procedures for this model of 6 predictors in the CLS as well as in Test sets 1 and 2.

Application to NHANES-III—To examine applicability of the prediction models in another cohort where bone lead levels are not available, we examined the associations between estimated bone lead levels and hypertension using data from the Third National Health and Nutrition Examination Survey (NHANES-III). The association between blood lead and hypertension has already been published elsewhere (29,30). Vupputuri et al. found that blood lead levels were significantly associated with hypertension only among both white and black women, borderline significantly associated with hypertension among black men, and not associated with hypertension among white men (30). We ran multiple logistic regression models with hypertension as an outcome and bone lead estimated using one of the three prediction models (full, reduced or 6-predictor model) as an exposure variable, adjusting for age, age², education, smoking status, pack-years of cigarette, BMI, hematocrit, alcohol consumption, physical activity, use of antihypertensive medication, and diagnosis of type-2 diabetes. We conducted stratified analyses by race and gender as done by Vupputuri et al (30). We also did stratified analyses by gender and age (<50 vs. ≥ 50 years) to see whether our models constructed from older males predict bone lead levels in younger population. We log-transformed blood lead to normalize the distribution but used non-log-transformed bone lead because those distributions were not very skewed and because bone lead levels include negative values. We illustrate how to use regression calibration techniques (31) to accommodate for prediction uncertainties in the imputed bone lead measurements in the logistic regression association models and obtain correct standard errors. All analyses were performed using the survey package in SAS (SAS Institute, Inc., Cary, North Carolina) to account for the complex sampling design.

Results

Development of prediction models

Table 1 shows the distributions of demographic and clinical characteristics and lead biomarkers among the baseline sample of 825 subjects in the NAS. The mean age was 67 years, and 29% and 53% of subjects were college graduates and white collar, respectively. There were no apparent differences in terms of descriptive statistics of participants in the training and test sets.

Figure 2 shows a visual description of LASSO model selection iterations in the tibia lead prediction model plotted against certain standard model assessment diagnostics like the AIC, BIC, Mallow's Cp and adjusted R^2 . These figures correspond to the fixed training set data. See Supplemental Figure 1 for the LASSO model selection for the patella lead. To arrive at a more robust set of predictors, we aggregated the model selection results over several training sets and collected the variables that appeared consistently over the iterations. Though the optimal dimension of the selected model varied across training sets, most frequently 15 predictors were included in tibia and 13 in the patella lead model besides 3 pre-selected predictors (age, education and job type) which are listed in Table 3.

Table 3 presents the regression models fitted to the complete case data in our original training set. In this training set 550 subjects had complete data on tibia lead and selected predictors, whereas 548 of these subjects had complete data on patella lead and the selected covariates. This difference in the number of complete case observations between tibia and patella lead models was due to missing data in different covariates used in the tibia lead and patella lead models. Blood lead, age, job type and education contributed almost all of the total variations explained by the two models. Blood lead had a stronger partial association with patella lead than tibia lead (partial $R^2=0.0924$ and 0.1253 for tibia and patella, respectively). Other variables that were statistically significant in both models were serum phosphorus ($P=0.001$ for tibia lead and $P=0.01$ for patella lead). Variables reaching significance in the tibia lead model only were serum uric acid ($P=0.005$), SBP ($P=0.04$) and current diagnosis of diabetes ($P=0.04$). Cumulative cigarette smoking as measured by pack-years appeared to have a stronger effect in the patella lead model ($P=0.09$) than the tibia lead model ($P=0.37$). Current diagnosis of hypertension ($P=0.01$) and high density lipoprotein (HDL) ($P=0.03$) also attained significance in the patella lead model. Finally, a set of 18 and 16 predictors were selected for tibia lead and patella lead, respectively. The different predictors included in each model are serum calcium, total cholesterol, and diagnosis of cancer for tibia lead; diagnosis of hypertension for patella lead. The R^2 for the patella lead model was slightly higher than that for the tibia lead model (0.335 vs. 0.316), and so was the adjusted R^2 value (0.313 vs. 0.290). The predictors listed in Table 3 were selected only if they were identified as important predictors more than 500 times out of 1000 different training sets by the LASSO procedure. Some variables with high p-values, such as serum creatinine ($p=0.85$) and serum calcium ($p=0.79$) for tibia lead and serum creatinine ($p=0.86$) and hematocrit ($p=0.82$) for patella lead, appeared not to be predictors of bone lead levels in the original training set, but they were identified as important predictors in other training sets (data not shown). Before we delve into the issue of prediction, we would like to mention that this model selection and fitting procedure provides guidance in terms of relative importance and statistical contribution of the variables determining the different lead burdens in body.

The fitted models were then used to predict the tibia bone lead levels of 188 subjects with completely observed data and, patella lead levels in 186 subjects with completely observed data in the first test set from NAS. Figure 3 presents validation of the fitted model in this test set. The correlations of the observed versus the predicted values were 0.454 for tibia lead and 0.524 for patella lead. The prediction was poor for the observations with higher observed values

of tibia and patella lead, which was expected due to lack of observations and increasing variability in those sparse domains. The bottom panel plots the errors in prediction (observed – predicted), again showing that a reasonable prediction accuracy was attained by using the fitted model. The mean absolute error for tibia lead was 7.9 (SD=7.4) $\mu\text{g/g}$, with median 6.5 (interquartile range (IQR)=6.2) $\mu\text{g/g}$, and that for patella lead was 10.2 (SD=9.4) $\mu\text{g/g}$ with median 7.7 (IQR=9.7) $\mu\text{g/g}$. These absolute errors do not appear to be much different compared to the measurement uncertainties of the instrument (median: 5.3 $\mu\text{g/g}$ for tibia lead; 6.9 $\mu\text{g/g}$ for patella lead).

We carried out a second validation on test set 2 as described in the methods section consisting of 118 subjects for tibia lead and 117 subjects for patella lead (see Supplemental Figure 2). We obtained slightly higher correlation coefficients for both tibia lead ($r=0.497$) and patella lead ($r=0.575$) than those in the test set 1. The mean absolute errors of prediction were also similar.

We also fit the models after dropping blood pressure and disease diagnoses (reduced model) which can be used when cancer, diabetes or blood pressure-related outcomes such as hypertension is examined (Table 4). Adjusted R^2 's were slightly reduced in both the tibia lead and patella lead models.

Prediction with fewer covariates and external validation

We also developed prediction models including NTx to account for individual differences in the bone resorption rate. NTx (1/NTx) and its interaction with blood lead were not significant predictors of both tibia and patella lead levels (data not shown), thus, including these variables did not improve the prediction models.

Since the models presented in Tables 3 and 4 contain measurements on many specific serum biomarkers and personal disease history variables which may not be available in other community based cohorts, we selected six basic variables which are likely to be routinely available in other cohorts and which account for a large contribution to the R^2 reported in Table 3. The results based on this model consisting of blood lead, age, education level, job type, smoking status and cumulative cigarette smoking (pack-years) are reported in Table 5. The validity of this model was tested on the two previously defined test sets constituted from the NAS. We found slightly better results in these models than in those with more covariates including serum measures and disease history (see Supplemental Figures 3 and 4). For example, the correlations of the observed versus the predicted values were 0.480 for tibia lead and 0.518 for patella lead in the test set 1. The mean absolute errors of prediction for tibia lead and patella lead were 7.5 (SD=7.3) $\mu\text{g/g}$ and 10.0 (SD=9.7) $\mu\text{g/g}$, respectively. Similar results were also found in the test set 2.

To examine the generalizability of the prediction equation to other cohorts beyond NAS, we used the prediction equation described in Table 5, to predict bone lead levels of 84 study participants in the CLS. Figure 4 presents the plot of observed vs predicted bone lead levels as well as the error of prediction. For CLS subjects, the correlation coefficients between observed and predicted values were 0.431 for tibia lead and 0.575 for patella lead. The mean absolute errors of prediction for the CLS subjects were 9.2 (SD=7.5) $\mu\text{g/g}$ with median 7.9 (IQR=9.2) $\mu\text{g/g}$ for tibia lead and 11.1 (SD=9.3) $\mu\text{g/g}$ with median 9.0 (IQR=13.6) $\mu\text{g/g}$ for patella lead levels. The absolute errors in this external validation seem to be somewhat higher compared to the measurement uncertainties in CLS (median 5.9 $\mu\text{g/g}$ for tibia lead; 7.5 $\mu\text{g/g}$ for patella lead) which could result from the different demographic characteristics of the NAS versus CLS population.

Application to NHANES-III

Blood lead and estimated bone lead levels among NHANES-III participants were presented in Table 6. The study sample included 12,500 subjects (5,870 men and 6,630 women) with a weighted mean $\pm$ SE age of 44.6 (0.46) years (range 20 – 90 years). Weighted mean $\pm$ SE blood lead level was 3.52 ± 0.10 $\mu\text{g/dL}$; higher for men compared with women and for blacks compared with whites. Estimated bone lead levels differed by the prediction models; the 6-predictor models had the highest bone lead levels (5.47 ± 0.35 $\mu\text{g/g}$ for tibia lead and 6.43 ± 0.52 $\mu\text{g/g}$ for patella lead) followed by the full models (3.81 ± 0.36 and 4.97 ± 0.52 $\mu\text{g/g}$) and the reduced models (3.32 ± 0.36 and 3.52 ± 0.53 $\mu\text{g/g}$). Estimated bone lead levels were also higher for men compared with women and for blacks compared with whites, but the weighted means among subjects younger than 50 years of age were all negative (except for black men) which resulted from negative values of the estimated bone lead levels for 20's and 30's. Subjects aged 50 years and older had substantially higher estimated bone lead levels from all three models.

Table 7 shows the associations between hypertension and blood lead and estimated bone lead levels among NHANES-III participants. After adjustment for potential confounders, a 1-SD increase in log-transformed blood lead (0.75 $\mu\text{g/dL}$) was associated with hypertension with an odds ratio (OR) of 1.12 (95% confidence interval (CI), 1.03, 1.23). As expected, the associations with estimated bone lead levels from the full models seem to be inflated (OR=3.70 (95% CI, 2.80, 4.89) for tibia lead (SD=15.2 $\mu\text{g/g}$); OR=4.95 (95% CI, 3.55, 6.90) for patella lead (SD=21.9 $\mu\text{g/g}$)). The associations with estimated bone lead levels from the reduced models appear to be much stable and stronger than those with blood lead (OR=2.06 (95% CI, 1.59, 2.67) for tibia lead (SD=15.1 $\mu\text{g/g}$); OR=1.69 (95% CI, 1.34, 2.14) for patella lead (SD=22.3 $\mu\text{g/g}$)). The 6-predictor models showed slightly larger effect sizes but similar significances to the blood lead model.

In stratified analyses by gender and race, the associations with blood lead were statistically significant only among women, borderline significant among black men and not significant among white men, which is consistent with the study by Vupputuri et al (30). The associations with bone lead markers from the reduced models were statistically significant for all 4 groups but the effect sizes among women seem to be overestimated. We found similar trends in the 6-predictor models.

In stratified analyses by gender and age (<50 vs. ≥ 50 years), blood lead was significantly associated with hypertension among men aged 50 years and older but among women younger than 50 years of age. The associations with bone lead markers from the reduced models among men were similar regardless of age groups, but those among women were larger in the older group. The associations with bone lead markers from the 6-predictor models among men were similar to those with blood lead, but among women the associations were larger in the older group.

Incorporating uncertainties in bone lead prediction in the association model

While the analysis reported above treats the predicted bone lead levels as truly observed values in the health outcome association model, incorporating prediction uncertainties in the standard error expression for the coefficients in the logistic regression model for hypertension in NHANES-III is necessary for achieving correct inference. One may adopt two methods to achieve this goal: (i) recall that bone lead is imputed from a regression model of the following nature: bone lead = linear function of predictors + ε , where ε is the normally distributed error with mean 0 and a standard deviation parameter that can be estimated from the square root of the mean squared error of the fitted linear model (11.26 for tibia, 16.70 for patella based on the fitted model in Table 3). One can generate M copies of the bone lead predictions by

generating M (typically 5-10) random normal variables from the estimated error distribution ($N(0, 11.26)$, say for tibia) and adding the errors to the point prediction. The errors can alternatively be generated from a log-normal distribution with parameters estimated from the observed error distribution in the training set to capture skewness in the bone-lead levels. Thus, one will have M datasets with M different imputed bone lead levels and each will give rise to a fitted logistic regression model and M sets of log OR estimates. The final estimate of the log OR and the corresponding variance derived from the M datasets can be obtained using the following equations (32).

$$\overline{\beta_j} = \sum_{m=1}^M \widehat{\beta}_j^{(m)} / k,$$

$$\text{var}(\overline{\beta_j}) = \sum [\widehat{\beta}_j^{(m)} - \overline{\beta_j}]^2 / (M - 1) + [(M+1)/M] \sum_{m=1}^M \text{var}(\widehat{\beta}_j^{(m)}) / M$$

Another approach is to adopt the regression calibration framework of Spiegelman et al (1997) to derive corrected large sample standard errors for the log OR. By applying this correction, 95% CI for the OR corresponding to 1-SD change in bone lead in the NHANES-III hypertension model (with bone lead as imputed via the full model) changes to (2.69, 5.12) for tibia lead; (3.38, 7.27) for patella lead. With imputation via the reduced model, 95% CIs are (1.57, 2.71) for tibia lead; (1.33, 2.16) for patella lead. For the 6-predictor model, they are (1.11, 1.83) for tibia lead and (1.09, 1.69) for patella lead. The point estimates of the OR remain unchanged.

Discussion

In this paper we derived prediction equations for bone lead levels in terms of blood lead and a set of covariates that may be measured in population-based cohorts. Our analysis served two purposes, first through model selection, we identified the important predictors, and then by validating on test sets and an external cohort and applying to a cohort where bone lead data are not available, we assessed the translatability of the coefficients to a set of independent observations. Based on the comparisons between measurement uncertainty and the absolute errors of prediction and the measurement errors computed from the regression models in NHANES-III, the prediction models including important determinants seem to predict bone lead levels reasonably. Although we could not assess how accurate estimated bone lead levels in NHANES-III were, the findings that estimated bone lead markers were more strongly associated with hypertension than blood lead provide evidence of the practical value of the present study, in particular, use of the reduced models in relation to hypertension, high blood pressure and disease diagnoses, and the potential use of the full models in relation to other health endpoints, such as cognitive declines. However, the associations of bone lead levels estimated from the 6-predictor models with hypertension did not appear to be different from those of blood lead. These findings suggest that biochemical blood parameters, such as serum phosphorus, uric acid, total and HDL cholesterols, may be critical contributors to distinguish predicted bone lead from blood lead and the models including these parameters may provide better estimations of bone lead levels than the 6-predictor models.

In this study, we verified the prediction equations using internal test datasets as well as an external cohort, the CLS. The correlations between the observed and the predicted bone lead levels were high in both validation tests, ranging 0.4-0.5 for tibia lead and 0.5-0.6 for patella lead. The CLS participants were all members of minority groups with 82% African American and 67% females, which are quite different characteristics from the NAS. Gender is an important determinant of bone lead levels. Kosnett et al. observed no significant difference in

tibia lead levels by gender among subjects less than 55 years age, but significantly higher tibia lead levels in males 55 years and older (19). This may be because the rate of bone remodeling increases in postmenopausal women: the bone resorption rate in postmenopausal women is accelerated, resulting in old bone being replaced relatively quickly by new bone matrix formed in a more recent lower-lead environment (14,33). Although the majority of the CLS participants were females, 75% of the females were aged less than 55 years (mean age 49 years), and therefore, the influence of postmenopause might be limited. Hence, the prediction equations may not be appropriate for populations including a number of postmenopausal women. In fact, much larger effects of estimated bone lead were observed among women aged 50 years and older in NHANES-III (Table 7). Because women showed a larger association between blood lead and hypertension in NHANES-III, it is unclear whether the larger effects observed among women aged 50 years and older indicate true associations or inflation, thus the association results should be taken as precursors for further epidemiological investigation. Also the equations should be used cautiously to predict bone lead levels for persons in a stage of high bone turnover other than postmenopause, such as pregnant women.

Race/ethnicity may also be an important factor to predict bone lead levels, but not included in the current prediction equations because the NAS participants were predominantly White. A recent study conducted in Baltimore found higher tibia lead levels in African Americans compared with Whites, after controlling for socio-economic status (SES) or other factors (17). The authors suggested that this difference might be due to the higher life-time cumulative exposure to more polluted environments, rather than differences in bone kinetics by race/ethnicity. Although the equations in this study predicted bone lead levels well even in the CLS, we cannot rule out a possibility that using the prediction equations may underestimate bone lead levels for African Americans or other members of minority groups.

Lead levels in blood generally reflect both recent exogenous exposures and the mobilization of lead from the skeleton back into the circulation (3,34). In the present study, a 1 $\mu\text{g}/\text{dL}$ difference in blood lead levels was associated with a 1.1 $\mu\text{g}/\text{g}$ change in tibia lead and a 1.9 $\mu\text{g}/\text{g}$ change in patella lead (Tables 3 and 4). Blood lead was the strongest predictor of bone lead levels which explained approximately 9% and 13% of the variances in tibia lead and patella lead, respectively. Because only one study reported blood lead as a determinant of tibia lead levels and that study log-transformed tibia lead by adding an offset value (19), we could not directly compare that study with ours. The study conducted with 101 suburban residents also found blood lead (log-transformed) as a significant predictor of tibia lead, but the association was not as strong as ours ($p=0.04$). This might be because of the small number of subjects with a wide age range (11 – 78 years). The current and our previous studies show a stronger association between blood lead and patella lead levels than between blood lead and tibia lead levels. Lead in trabecular bone, such as patella bone, is more readily released than lead in cortical bone, and therefore, this higher correlation between blood lead and patella lead led to better validation results.

The equations may be more feasible for older populations not-occupationally exposed to lead. Predicted bone lead levels for older age groups seem to be comparable to the measured bone lead levels in other cohorts. Measured tibia lead levels for white men of the NAS (mean age $\pm$ SD: 67.2 ± 7.3 , see Table 1), white women of the Nurses' Health Study (NHS, mean age $\pm$ SD: 58.7 ± 7.2) (35), and blacks (mean age $\pm$ SD: 59.7 ± 6.2) and whites (mean age $\pm$ SD: 59.1 ± 5.7) of the Baltimore Memory Study (36) were 21.9, 13.3, 21.5 and 16.7 $\mu\text{g}/\text{g}$, respectively. The estimated tibia lead levels in the older group of the NHANES-III were 19.2 $\mu\text{g}/\text{g}$ for white men, 16.5 $\mu\text{g}/\text{g}$ for white women, 24.7 $\mu\text{g}/\text{g}$ for black men and 20.4 $\mu\text{g}/\text{g}$ for black women (Table 6). Measured patella lead levels in NAS (white men) and NHS (white women) were 31.1 and 17.3 $\mu\text{g}/\text{g}$, respectively, and the estimated patella lead levels for the corresponding groups in NHANES-III were 27.5 and 23.2 $\mu\text{g}/\text{g}$, respectively.

Previous studies have consistently reported that age is the predominant predictor of bone lead levels (13-16,18,19). A one-year increase in age was positively associated with 0.28 to 0.63 $\mu\text{g/g}$ increases in tibia lead and 0.24 to 0.72 $\mu\text{g/g}$ increases in patella lead. In our study, the corresponding regression coefficients were 0.54 to 0.60 $\mu\text{g/g}$ for tibia lead and 0.86 to 0.91 $\mu\text{g/g}$ for patella lead. Age accounted for about 7 to 10% of the variance in bone lead levels, one third of the model R^2 observed (Tables 3-5). Studies conducted by Kosnett et al. (19) and Lin et al. (13) suggested that age-related bone lead levels differ by age groups: tibia lead levels increased at a greater rate among persons aged 55 years and older, as compared with those < 55 years. This might reflect the age-related difference in bone turnover (26) or the birth cohort effect that exposure levels to environmental lead are different over time (37). Because our prediction models were derived from an older age cohort (the range of age 48-94 years), the bone lead levels for young participants (less than 50 years of age) in NHANES-III did not appear to be adequately predicted: estimated bone lead levels in this age group were negative. Nonetheless, this does not necessarily mean that examination of the effect of estimated bone lead in younger population is inadequate; we could detect positive associations between estimated bone lead and hypertension in this age group (Table 7).

Education, job-type (white collar or not) and cigarette smoke were identified as strong determinants of bone lead levels. Educational attainment and job-type are proxies of SES and low SES is a well-known predictor of bone lead levels (3,24). In studies with populations with occupational lead exposure, exposure duration was another important determinant of bone lead levels (15,16), which could not be examined in the present study. Because details of occupation history are unlikely collected in community-based studies, even the crude white or blue collar occupation classification improves the prediction models. Smoking can also predict bone lead levels well because lead exposure may occur through direct lead intake from tobacco, increased hand-to-mouth activity, and enhanced permeability of the smoke-exposed respiratory tract (38,39).

Biological data are often right-skewed and non-normally distributed and distributions of bone lead levels in the present study were not normal. In general, the assumption of normality is particularly important for hypothesis testing and construction of confidence intervals. However, the primary objective of our study is to provide point predictions for bone lead levels, not to do hypothesis testing with bone lead as response, and thus, these assumptions are not critical for our prediction rule. The least square fitting may be evaluated as an optimization technique to yield the best fitted surface irrespective of the distributional assumption of a standard normal model. In addition, negative bone lead values are sometimes produced when the true value is close to zero (22) and in fact, there were negative bone lead levels in this cohort which precludes log-transformation. Therefore, we used non-transformed bone lead levels as the target outcome for prediction. Also, while correcting for prediction uncertainty in the second-step logistic regression association models with predicted bone lead as a risk factor, normality assumption is not required (40).

Several limitations should be considered in this study. Many other factors that were not considered in the prediction equations may contribute to bone lead prediction, such as diet and genetic polymorphisms. We considered NTx, a marker of bone resorption, but it was not identified as an important predictor of bone lead levels. In the NAS, our group previously found that dietary intakes of vitamin D, phosphorus and calcium were associated with bone lead levels (12). However, those dietary variables were not identified as important predictors in the present study, probably because of high correlations with serum phosphorus and calcium which might be better markers than dietary measures. We included those serum markers because they are routinely collected from a simple blood test, but not all population studies collect dietary intakes. In addition, our group has found that bone lead levels differ by genetic polymorphisms,

such as δ -aminolevulinic acid dehydratase and hemochromatosis genes (41,42). Including these genetic polymorphisms would provide better prediction models.

We presented the results for one specific training set and test sets, while the coefficients may change in the next training set selected. One could average all fitted equations over a large number of training sets and the prediction accuracy over test sets, to account for this uncertainty. However, we refrained from that approach as description, interpretation of model coefficients, as well as visualization of the prediction plots would become cumbersome in that context.

The results indicate that in fact the prediction equation fitted in NAS is translatable to other cohorts which may not exactly be similar to the NAS cohort in demographic and socio-economic characteristics. The consistency of the reported values across test sets and different models are indeed reassuring in the sense that in spite of sampling variability, and differences in the study population, one can ensure approximately 0.4-0.5 correlation between observed and predicted bone lead measurements.

If one were to impute the bone lead levels in other studies based on this equation, it is important to include a measure of the uncertainty associated with this prediction and to adjust the variance estimates of the obtained effect sizes accordingly. A naïve approach of assuming the predicted bone lead levels as true values will underestimate the actual variances and inflate Type I error properties of the tests based on imputed data. Prediction uncertainty measures and prediction intervals for the measurements can be obtained, using the estimated error variability from the NAS study.

In summary, this study suggests that the prediction equations developed in the NAS may be used to predict bone lead levels in other community-based cohorts with reasonable accuracy based on blood lead levels and other standard covariates where bone lead measurements are not available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Disclosure: This study was funded by the National Institute of Environmental Health Sciences (NIEHS) R01-ES05257, ES10798, and P30-ES00002. S.K.P was supported by the National Institute of Environmental Health Sciences (NIEHS) K01-ES016587-01A1.

All Sources of Support: This research was supported primarily by the National Institute of Environmental Health Sciences (NIEHS) R01-ES05257, ES10798, and P30-ES00002. S.K.P was supported by the National Institute of Environmental Health Sciences (NIEHS) K01-ES016587-01A1. The VA Normative Aging Study is supported by the Cooperative Studies Program/Epidemiology Research and Information Center of the US Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), Boston, Massachusetts. The authors declare they have no competing financial interests.

References

1. Pirkle JL, Brody DJ, Gunter EW, et al. The decline in blood lead levels in the United States. The National Health and Nutrition Examination Surveys (NHANES). JAMA 1994;272:284–291. [PubMed: 8028141]
2. Muntner P, Menke A, DeSalvo KB, Rabito FA, Batuman V. Continued decline in blood lead levels among adults in the United States: the National Health and Nutrition Examination Surveys. Arch Intern Med 2005;165:2155–2161. [PubMed: 16217007]

3. Hu H, Shih R, Rothenberg S, Schwartz BS. The epidemiology of lead toxicity in adults: measuring dose and consideration of other methodologic issues. *Environ Health Perspect* 2007;115:455–462. [PubMed: 17431499]
4. Hu H, Rabinowitz M, Smith D. Bone lead as a biological marker in epidemiologic studies of chronic toxicity: conceptual paradigms. *Environ Health Perspect* 1998;106:1–8. [PubMed: 9417769]
5. Rabinowitz MB. Toxicokinetics of bone lead. *Environ Health Perspect* 1991;91:33–37. [PubMed: 2040248]
6. Leggett RW. An age-specific kinetic model of lead metabolism in humans. *Environ Health Perspect* 1993;101:598–616. [PubMed: 8143593]
7. Saltzman BE, Gross SB, Yeager DW, Meiners BG, Gartside PS. Total body burdens and tissue concentrations of lead, cadmium, copper, zinc, and ash in 55 human cadavers. *Environ Res* 1990;52:126–145. [PubMed: 2394204]
8. Barry PS, Mossman DB. Lead concentrations in human tissues. *Br J Ind Med* 1970;27:339–351. [PubMed: 5488693]
9. Hu H, Aro A, Payton M, et al. The relationship of bone and blood lead to hypertension. The Normative Aging Study. *JAMA* 1996;275:1171–1176. [PubMed: 8609684]
10. Schaumberg DA, Mendes F, Balaram M, Dana MR, Sparrow D, Hu H. Accumulated lead exposure and risk of age-related cataract in men. *JAMA* 2004;292:2750–2754. [PubMed: 15585735]
11. Weisskopf MG, Wright RO, Schwartz J, et al. Cumulative lead exposure and prospective change in cognition among elderly men: the VA Normative Aging Study. *Am J Epidemiol* 2004;160:1184–1193. [PubMed: 15583371]
12. Cheng Y, Willett WC, Schwartz J, Sparrow D, Weiss S, Hu H. Relation of nutrition to bone lead and blood lead levels in middle-aged to elderly men. The Normative Aging Study. *Am J Epidemiol* 1998;147:1162–1174. [PubMed: 9645795]
13. Lin C, Kim R, Tsaih SW, Sparrow D, Hu H. Determinants of bone and blood lead levels among minorities living in the Boston area. *Environ Health Perspect* 2004;112:1147–1151. [PubMed: 15289158]
14. Korrick SA, Schwartz J, Tsaih SW, et al. Correlates of bone and blood lead levels among middle-aged and elderly women. *Am J Epidemiol* 2002;156:335–343. [PubMed: 12181103]
15. Schwartz BS, Stewart WF, Todd AC, Links JM. Predictors of dimercaptosuccinic acid chelatable lead and tibial lead in former organolead manufacturing workers. *Occup Environ Med* 1999;56:22–29. [PubMed: 10341742]
16. Todd AC, Lee BK, Lee GS, Ahn KD, Moshier EL, Schwartz BS. Predictors of DMSA chelatable lead, tibial lead, and blood lead in 802 Korean lead workers. *Occup Environ Med* 2001;58:73–80. [PubMed: 11160984]
17. Theppeang K, Glass TA, Bandeen-Roche K, Todd AC, Rohde CA, Schwartz BS. Gender and race/ethnicity differences in lead dose biomarkers. *Am J Public Health* 2008;98:1248–1255. [PubMed: 18511728]
18. Hu H, Payton M, Korrick S, et al. Determinants of bone and blood lead levels among community-exposed middle-aged to elderly men. The normative aging study. *Am J Epidemiol* 1996;144:749–759. [PubMed: 8857824]
19. Kosnett MJ, Becker CE, Osterloh JD, Kelly TJ, Pasta DJ. Factors influencing bone lead concentration in a suburban community assessed by noninvasive K x-ray fluorescence. *Jama* 1994;271:197–203. [PubMed: 8277545]
20. Hu H, Milder FL, Burger DE. X-ray fluorescence measurements of lead burden in subjects with low-level community lead exposure. *Arch Environ Health* 1990;45:335–341. [PubMed: 2270952]
21. Burger DE, Milder FL, Morsillo PR, Adams BB, Hu H. Automated bone lead analysis by K-x-ray fluorescence for the clinical environment. *Basic Life Sci* 1990;55:287–292. [PubMed: 2088281]
22. Kim R, Aro A, Rotnitzky A, Amarasiriwardena C, Hu H. K x-ray fluorescence measurements of bone lead concentration: the analysis of low-level data. *Phys Med Biol* 1995;40:1475–1485. [PubMed: 8532760]
23. Elmarsafawy SF, Tsaih SW, Korrick S, et al. Occupational determinants of bone and blood lead levels in middle aged and elderly men from the general community: the Normative Aging Study. *Am J Ind Med* 2002;42:38–49. [PubMed: 12111689]

24. Elreedy S, Krieger N, Ryan PB, Sparrow D, Weiss ST, Hu H. Relations between individual and neighborhood-based measures of socioeconomic position and bone lead concentrations among community-exposed men: the Normative Aging Study. *Am J Epidemiol* 1999;150:129–141. [PubMed: 10412957]
25. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J Royal Statist Soc B* 1996;58:267–288.
26. Vig EK, Hu H. Lead toxicity in older adults. *J Am Geriatr Soc* 2000;48:1501–1506. [PubMed: 11083332]
27. Hanson DA, Weis MA, Bollen AM, Maslan SL, Singer FR, Eyre DR. A specific immunoassay for monitoring human bone resorption: quantitation of type I collagen cross-linked N-telopeptides in urine. *J Bone Miner Res* 1992;7:1251–1258. [PubMed: 1466251]
28. Tsaih SW, Korrick S, Schwartz J, et al. Influence of bone resorption on the mobilization of lead from bone among middle-aged and elderly men: the Normative Aging Study. *Environ Health Perspect* 2001;109:995–999. [PubMed: 11675263]
29. Den Hond E, Nawrot T, Staessen JA. The relationship between blood pressure and blood lead in NHANES III. National Health and Nutritional Examination Survey. *J Hum Hypertens* 2002;16:563–568. [PubMed: 12149662]
30. Vupputuri S, He J, Muntner P, Bazzano LA, Whelton PK, Batuman V. Blood lead level is associated with elevated blood pressure in blacks. *Hypertension* 2003;41:463–468. [PubMed: 12623944]
31. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr* 1997;65:1179S–1186S. [PubMed: 9094918]
32. Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. New Jersey: Wiley Inter Science; 2002.
33. Nash D, Magder LS, Sherwin R, Rubin RJ, Silbergeld EK. Bone density-related predictors of blood lead level among peri- and postmenopausal women in the United States: The Third National Health and Nutrition Examination Survey, 1988–1994. *Am J Epidemiol* 2004;160:901–911. [PubMed: 15496543]
34. Hu H. Bone lead as a new biologic marker of lead dose: recent findings and implications for public health. *Environ Health Perspect* 1998;106 4:961–967. [PubMed: 9703479]
35. Korrick SA, Hunter DJ, Rotnitzky A, Hu H, Speizer FE. Lead and hypertension in a sample of middle-aged women. *Am J Public Health* 1999;89:330–335. [PubMed: 10076481]
36. Martin D, Glass TA, Bandeen-Roche K, Todd AC, Shi W, Schwartz BS. Association of blood lead and tibia lead with blood pressure and hypertension in a community sample of older adults. *Am J Epidemiol* 2006;163:467–478. [PubMed: 16421242]
37. Kim R, Landrigan C, Mossman P, Sparrow D, Hu H. Age and secular trends in bone lead levels in middle-aged and elderly men: three-year longitudinal follow-up in the Normative Aging Study. *Am J Epidemiol* 1997;146:586–591. [PubMed: 9326437]
38. Willers S, Schutz A, Attewell R, Skerfving S. Relation between lead and cadmium in blood and the involuntary smoking of children. *Scand J Work Environ Health* 1988;14:385–389. [PubMed: 3212414]
39. Wietlisbach V, Rickenbach M, Berode M, Guillemin M. Time trend and determinants of blood lead levels in a Swiss population over a transition period (1984–1993) from leaded to unleaded gasoline use. *Environ Res* 1995;68:82–90. [PubMed: 7601076]
40. Carroll RJ, Stefanski LA. Approximate quasi-likelihood estimation in models with surrogate predictors. *J Amer Statist Assoc* 1990;85:652–663.
41. Hu H, Wu MT, Cheng Y, Sparrow D, Weiss S, Kelsey K. The delta-aminolevulinic acid dehydratase (ALAD) polymorphism and bone and blood lead levels in community-exposed men: the Normative Aging Study. *Environ Health Perspect* 2001;109:827–832. [PubMed: 11564619]
42. Wright RO, Silverman EK, Schwartz J, et al. Association between hemochromatosis genotype and lead exposure among elderly men: the Normative Aging Study. *Environ Health Perspect* 2004;112:746–750. [PubMed: 15121519]

Figure 1.
Procedures of the model development and validation.

Figure 2.

LASSO model selection for the tibia bone lead model. We performed the LASSO with the residuals corresponding to a linear regression model of tibia bone lead on the three pre-selected variables, age, education, and job-type (fitted on the training set) as response and using the remaining set of 25 potential predictors (smoking status was represented by two dummy variables (former and current smokers), thus, leading to a total of 26 LASSO iterations). The arrows on the graph point to the models selected by the LASSO operator. AIC, Akaike Information Criterion; BIC, Bayes information criterion; Cp, Mallows' Cp.

Figure 3.

Validation of the prediction models (see Table 3) in test set 1. The numbers of subjects are 188 for tibia lead and 186 for patella lead. Figures in the upper panel show relationships of the observed versus the predicted values. The lines drawn on the plots indicate the equiangular observed=predicted line. The correlations of the observed versus the predicted are 0.454 for tibia lead and 0.524 for patella lead. The lower panels are scatterplots of the errors in prediction (observed – predicted).

Figure 4.

Validation of the prediction models (see Table 5) in the Community Lead Study. The number of subjects is 84 for each tibia lead or patella lead. Figures in the upper panel show relationships of the observed versus the predicted values. The lines drawn on the plots indicate the equiangular observed=predicted line. The correlations of the observed versus the predicted are 0.431 for tibia lead and 0.575 for patella lead. The lower panels are scatterplots of the errors in prediction (observed – predicted).

Table 1

Demographic and clinical characteristics and lead biomarkers among participants in the Normative Aging Study (N = 825).

	Mean $\pm$ SD	Range
Age (years)	67.2 $\pm$ 7.3	48 – 94
Body mass index (kg/m ²)	27.9 $\pm$ 3.9	16.7 – 44.9
Height (cm)	173.2 $\pm$ 6.7	150.2 – 195.6
Weight (kg)	83.9 $\pm$ 13.3	52.7 – 134.8
Hip circumference (cm)	101.9 $\pm$ 7.4	83.2 – 140.2
Waist circumference (cm)	98.5 $\pm$ 9.8	70.1 $\pm$ 140.6
Systolic blood pressure (mm Hg)	135.8 $\pm$ 17.3	91 – 215
Diastolic blood pressure (mm Hg)	81.4 $\pm$ 9.6	51 – 119
Serum creatinine (mg/dL)	1.2 $\pm$ 0.36	0.6 – 9.0
Serum uric acid (mg/dL)	6.4 $\pm$ 1.3	2.9 – 10.6
Serum calcium (mg/dL)	9.5 $\pm$ 0.42	8.3 – 11.3
Serum phosphorus (mg/dL)	3.1 $\pm$ 0.48	1.8 – 4.6
Serum triglyceride (mg/dL)	154.5 $\pm$ 95.3	24 – 795
Total cholesterol (mg/dL)	227.2 $\pm$ 37.4	80 – 438
High density lipoprotein (mg/dL)	48.2 $\pm$ 12.5	15 – 131
Hematocrit (%)	43.8 $\pm$ 3.1	29 – 52
Alcohol intake (g/day)	13.0 $\pm$ 17.9	0 – 104.1
Physical activity (kcal/week)	2,471 $\pm$ 2,041	196 – 14,896
Cumulative cigarette (pack-years)	21.7 $\pm$ 25.3	0 – 136
Number of years smoked cigarettes (years)	17.1 $\pm$ 16.1	0 – 58
Lead biomarkers		
Blood lead (μ g/dL)	6.0 $\pm$ 4.0	0 – 35
Tibia lead (μ g/g)	21.9 $\pm$ 13.6	-5 – 126
Measurement uncertainty (μ g/g)	5.6 $\pm$ 1.1	0.5 – 10
Patella lead (μ g/g)	31.1 $\pm$ 19.5	-10 – 165
Measurement uncertainty (μ g/g)	7.0 $\pm$ 1.4	1.0 – 14.3
N (%)		
Smoking status		
Never	247 (29.9)	
Former	510 (61.8)	
Current	68 (8.2)	
Education		
High school dropout	85 (10.3)	
High school graduate and some college	503 (61.0)	
$\geq$ 4 years of college	237 (28.7)	
White collar	435 (52.9)	
Diagnosis of cancer	119 (14.5)	
Diagnosis of diabetes	55 (6.7)	

	Mean $\pm$ SD	Range
Diagnosis of hypertension	340 (41.4)	
Diagnosis of ischemic heart disease	134 (16.3)	

Table 2

Characteristics of the CLS population used in the validation study.

	Mean $\pm$ SD or N (%)
Tibia lead, $\mu\text{g/g}$	11.9 $\pm$ 11.0
Measurement uncertainty ($\mu\text{g/g}$)	6.0 $\pm$ 1.4
Patella lead, $\mu\text{g/g}$	14.2 $\pm$ 15.3
Measurement uncertainty ($\mu\text{g/g}$)	7.8 $\pm$ 2.1
Blood lead, $\mu\text{g/dL}$	3.0 $\pm$ 2.2
Age, years	49.9 $\pm$ 10.1
Cumulative smoking, pack-years	7.6 $\pm$ 11.3
Smoking status	
Never	28 (33.3)
Former	26 (31.0)
Current	30 (35.7)
Education	
High school dropout	14 (16.7)
High school graduate and some college	27 (32.1)
≥ 4 years of college	43 (51.2)
Female	56 (66.7)
African American	69 (82.1)
White collar	57 (67.9)

Table 3

Regression coefficients of the prediction models in the training set determined by the LASSO model selection procedures (FULL model).

Variable	Tibia lead (n=550)				Patella lead (n=548)			
	β	SE	P	Partial R ^{2a}	β	SE	P	Partial R ^{2a}
Intercept	-16.558	16.504	0.32		-29.827	19.140	0.12	
Blood lead	1.064	0.126	<.0001	0.0924	1.854	0.186	<.00001	0.1253
Age	0.539	0.080	<.0001	0.0585	0.861	0.118	<.0001	0.0669
Education ^b								
High school diploma	-4.070	1.745	0.02	0.0070	-4.249	2.574	0.10	0.0034
≥ 4 years of college	-7.173	2.093	0.0007	0.0152	-8.634	3.101	0.006	0.0097
White collar	-2.962	1.180	0.01	0.0082	-4.020	1.758	0.02	0.0066
Cumulative cigarette smoking (pack-years)	0.023	0.026	0.37	0.0010	0.066	0.038	0.09	0.0037
Smoking status ^c								
Former smoker	1.652	1.316	0.21	0.0020	1.753	1.947	0.37	0.0010
Current smoker	2.416	2.494	0.33	0.0012	3.780	3.692	0.31	0.0013
Serum phosphorus	-3.815	1.169	0.001	0.0138	-4.408	1.687	0.01	0.0086
Serum uric acid	1.197	0.428	0.005	0.0101	1.173	0.619	0.06	0.0045
Serum calcium	-0.345	1.317	0.79	9E-05	--	--	--	--
Serum creatinine	-0.238	1.283	0.85	4E-05	0.337	1.895	0.86	4E-5
Total cholesterol	0.019	0.014	0.18	0.0024	--	--	--	--
Total HDL	0.061	0.042	0.14	0.0028	0.134	0.060	0.03	0.0062
Hematocrit	-0.188	0.166	0.26	0.0017	-0.053	0.240	0.82	6E-5
Body mass index	0.091	0.147	0.54	0.0005	0.315	0.214	0.14	0.0027
Systolic blood pressure	0.077	0.037	0.04	0.0056	-0.029	0.058	0.62	0.0003
Diastolic blood pressure	-0.062	0.069	0.37	0.0010	-0.112	0.103	0.28	0.0015
Diagnosis of cancer	1.306	1.359	0.34	0.0012	--	--	--	--
Diagnosis of diabetes	4.493	2.188	0.04	0.0055	3.468	3.229	0.28	0.0015
Diagnosis of hypertension	--	--	--		4.737	1.794	0.01	0.0088
AIC	2703.5				3121.9			
BIC	2707.1				3125.2			
R ²	0.316				0.335			

Variable	Tibia lead (n=550)				Patella lead (n=548)			
	β	SE	P	Partial R ^{2a}	β	SE	P	Partial R ^{2a}
Adjusted R ²	0.290				0.313			

^aPartial R² was calculated based on type II sum of squares.

^bReference group is high school dropout.

^cReference group is never smokers.

Table 4

Regression coefficients of the prediction models after excluding blood pressure and disease diagnoses from Table 3 (REDUCED model).

Variable	Tibia lead (n=550)				Patella lead (n=548)			
	β	SE	P	Partial R ^{2a}	β	SE	P	Partial R ^{2a}
Intercept	-12.603	16.109	0.43		-42.927	18.003	0.02	
Blood lead	1.065	0.125	<.0001	0.0944	1.851	0.185	<.0001	0.1272
Age	0.604	0.075	<.0001	0.0847	0.911	0.110	<.0001	0.0864
Education ^b								
High school diploma	-4.244	1.751	0.02	0.0077	-4.125	2.577	0.11	0.0033
≥ 4 years of college	-7.362	2.094	0.0005	0.0162	-8.295	3.090	0.008	0.0091
White collar	-2.994	1.185	0.01	0.0084	-4.374	1.755	0.01	0.0079
Cumulative cigarette smoking (pack-years)	0.027	0.026	0.30	0.0014	0.073	0.038	0.06	0.0046
Smoking status ^c								
Former smoker	1.808	1.321	0.17	0.0025	1.805	1.953	0.36	0.0011
Current smoker	2.044	2.492	0.41	0.0009	3.648	3.685	0.32	0.0012
Serum phosphorus	-3.668	1.170	0.002	0.0129	-4.178	1.683	0.01	0.0078
Serum uric acid	1.048	0.422	0.01	0.0081	1.108	0.609	0.07	0.0042
Serum calcium	-0.403	1.318	0.76	0.0001	--	--	--	--
Serum creatinine	-0.223	1.279	0.86	4E-05	0.226	1.890	0.91	2E-05
Total cholesterol	0.019	0.014	0.18	0.0023	--	--	--	--
Total HDL	0.060	0.042	0.15	0.0027	0.124	0.060	0.04	0.0053
Hematocrit	-0.257	0.164	0.12	0.0032	-0.106	0.237	0.65	0.0003
Body mass index	0.156	0.143	0.27	0.0016	0.362	0.208	0.08	0.0039
AIC	2705.8				3122.8			
BIC	2708.9				3125.7			
R ²	0.303				0.324			
Adjusted R ²	0.282				0.307			

^aPartial R² was calculated based on type II sum of squares.

^bReference group is high school dropout.

^cReference group is never smokers.

Table 5

Regression coefficients of the prediction models with basic 6 covariates in the training set (6-PREDICTOR model).

Variable	Tibia lead (n=550)				Patella lead (n=548)			
	β	SE	P	Partial R ²	β	SE	P	Partial R ²
Intercept	-20.271	5.341	0.0002		-34.486	7.849	<.0001	
Blood lead	1.035	0.125	<.0001	0.0918	1.844	0.184	<.0001	0.1291
Age	0.591	0.071	<.0001	0.0919	0.876	0.105	<.0001	0.0896
Education ^b								
High school diploma	-3.651	1.750	0.04	0.0059	-3.988	2.572	0.12	0.0031
≥ 4 years of college	-7.046	2.089	0.0008	0.0153	-8.190	3.072	0.008	0.0092
White collar	-3.211	1.184	0.007	0.0099	-4.822	1.744	0.006	0.0098
Cumulative cigarette smoking (pack-years)	0.035	0.026	0.17	0.0025	0.082	0.038	0.03	0.0059
Smoking status ^c								
Former smoker	1.797	1.335	0.18	0.0024	1.775	1.963	0.37	0.0011
Current smoker	0.049	2.480	0.98	5E-07	1.359	3.645	0.71	0.0002
AIC	2713.2				3125.1			
BIC	2715.4				3127.4			
R square	0.272				0.306			
Adjusted R square	0.262				0.296			

^aPartial R² was calculated based on type II sum of squares.

^bReference group is high school dropout.

^cReference group is never smokers.

Table 6

Distributions of blood lead and estimated bone lead levels among NHANES-III participants.

Unweighted, No.	Percentage of sample	Blood lead ($\mu\text{g/dL}$)	Weighted (mean $\pm$ SE)					
			Estimated tibia lead ($\mu\text{g/g}$)			Estimated patella lead ($\mu\text{g/g}$)		
			Full	Reduced	6-predictor	Full	Reduced	6-predictor
All	12,500	3.52 $\pm$ 0.10	3.81 $\pm$ 0.36	3.32 $\pm$ 0.36	5.47 $\pm$ 0.35	4.97 $\pm$ 0.52	3.52 $\pm$ 0.53	6.43 $\pm$ 0.52
By gender, race and age								
White men								
< 50 yrs	2,130	4.02 $\pm$ 0.16	-1.78 $\pm$ 0.33	-2.78 $\pm$ 0.32	-0.23 $\pm$ 0.30	-3.47 $\pm$ 0.47	-4.95 $\pm$ 0.47	-1.72 $\pm$ 0.47
$\geq$ 50 yrs	2,152	4.92 $\pm$ 0.18	19.2 $\pm$ 0.31	18.8 $\pm$ 0.32	20.0 $\pm$ 0.31	27.5 $\pm$ 0.47	27.0 $\pm$ 0.47	28.6 $\pm$ 0.48
Black men								
< 50 yrs	1,048	4.55 $\pm$ 0.15	-0.37 $\pm$ 0.39	-1.55 $\pm$ 0.38	0.57 $\pm$ 0.33	-1.51 $\pm$ 0.60	-2.98 $\pm$ 0.59	-0.52 $\pm$ 0.52
$\geq$ 50 yrs	540	7.57 $\pm$ 0.22	24.7 $\pm$ 0.45	24.0 $\pm$ 0.42	23.9 $\pm$ 0.35	35.2 $\pm$ 0.64	34.4 $\pm$ 0.65	34.6 $\pm$ 0.56
White women								
< 50 yrs	2,429	2.09 $\pm$ 0.07	-6.13 $\pm$ 0.20	-6.27 $\pm$ 0.21	-3.33 $\pm$ 0.15	-9.06 $\pm$ 0.30	-11.1 $\pm$ 0.29	-7.01 $\pm$ 0.22
$\geq$ 50 yrs	2,180	3.53 $\pm$ 0.12	16.5 $\pm$ 0.33	16.3 $\pm$ 0.30	17.7 $\pm$ 0.31	23.2 $\pm$ 0.45	22.2 $\pm$ 0.43	24.3 $\pm$ 0.47
Black women								
< 50 yrs	1,409	2.52 $\pm$ 0.09	-4.80 $\pm$ 0.33	-5.02 $\pm$ 0.34	-2.99 $\pm$ 0.28	-7.31 $\pm$ 0.48	-9.33 $\pm$ 0.49	-6.51 $\pm$ 0.42
$\geq$ 50 yrs	612	4.49 $\pm$ 0.16	20.4 $\pm$ 0.49	19.8 $\pm$ 0.43	19.6 $\pm$ 0.35	28.5 $\pm$ 0.68	26.9 $\pm$ 0.61	27.0 $\pm$ 0.51

Adjusted odds ratios and 95% confidence intervals of hypertension associated with a 1-SD higher level of lead markers among NHANES-III participants.

Table 7

SD	Weighted Prevalence	Log BLL	Estimated tibia lead			Estimated patella lead		
			Full		6-predictor	Full		6-predictor
			0.75 µg/dL	15.2 µg/g		15.1 µg/g	21.9 µg/g	
All subjects ^a	31.2%	1.12 (1.03, 1.23)	3.70 (2.80, 4.89)	2.06 (1.59, 2.67)	1.43 (1.11, 1.83)	4.95 (3.55, 6.90)	1.69 (1.34, 2.14)	1.36 (1.10, 1.69)
By gender and race groups ^b								
Men								
White	31.6%	1.06 (0.92, 1.22)	2.79 (1.98, 3.93)	1.79 (1.32, 2.43)	1.16 (0.82, 1.64)	3.01 (1.98, 4.56)	1.46 (1.09, 1.95)	1.13 (0.83, 1.54)
Black	34.9%	1.17 (0.98, 1.38)	2.47 (1.67, 3.66)	1.64 (1.11, 2.41)	1.32 (0.93, 1.87)	2.27 (1.59, 3.23)	1.38 (1.03, 1.85)	1.30 (0.96, 1.76)
Women								
White	29.7%	1.16 (1.04, 1.29)	7.18 (4.20, 12.29)	2.85 (1.78, 4.58)	2.08 (1.16, 3.74)	17.29 (9.87, 30.31)	2.31 (1.46, 3.64)	2.00 (1.19, 3.36)
Black	36.6%	1.19 (1.04, 1.38)	8.05 (5.19, 12.48)	4.05 (2.59, 6.35)	3.86 (2.10, 7.09)	12.43 (6.35, 24.35)	3.45 (2.09, 5.71)	3.38 (1.94, 5.86)
By gender and age groups ^c								
Men								
< 50 yrs	21.4%	0.98 (0.80, 1.22)	2.14 (1.38, 3.31)	1.80 (1.16, 2.79)	0.97 (0.60, 1.59)	2.84 (1.69, 4.76)	1.40 (0.93, 2.09)	0.96 (0.63, 1.47)
≥ 50 yrs	53.2%	1.20 (1.02, 1.41)	3.96 (2.58, 6.07)	1.71 (1.17, 2.51)	1.46 (0.96, 2.22)	3.07 (2.07, 4.55)	1.53 (1.07, 2.18)	1.40 (0.97, 2.02)
Women								
< 50 yrs	15.8%	1.23 (1.04, 1.46)	3.90 (2.15, 7.08)	2.15 (1.12, 4.11)	2.08 (1.14, 3.81)	26.3 (12.6, 54.9)	2.00 (1.11, 3.60)	1.95 (1.16, 3.30)
≥ 50 yrs	56.5%	1.09 (0.94, 1.26)	11.56 (7.64, 17.48)	3.53 (2.33, 5.34)	2.40 (1.32, 4.35)	10.9 (7.20, 16.50)	2.94 (1.92, 4.50)	2.16 (1.29, 3.63)

^aModels for all subjects were adjusted for age (years), age² (years²), gender, race (white/black), education (less than high school/high school diploma/college or higher), smoking status (never/former/current), cumulative smoking (pack-years), body mass index (kg/m²), hemocrit (%), alcohol intake (yes/no), physical activity (yes/no), use of antihypertensive medication (yes/no), diagnosis of type-2 diabetes (yes/no).

^bStratified models by gender and race have the same covariates included in the models for all subjects except gender and race.

^c Stratified models by gender and age have the same covariates included in the models for all subjects except gender.