A prediction model of Nephrolithiasis Risk: A population-based cohort study in Korea

David Mukasa[1] , Joohon Sung[1,2]

[1]Complex Disease and Genome Epidemiology Branch, Department of Public Health, Graduate School of Public Health, Seoul National University, Seoul,
[2]Department of Epidemiology, Graduate School of Public Health and Institute of Health and Environment, Seoul National University, Seoul, Korea

Purpose: Well-validated risk prediction models help to stratify individuals on the basis of their disease risks and to guide health care professionals in decision-making. The incidence of nephrolithiasis has been increasing in Korea. Racial differences in the distribution of and risk for nephrolithiasis have been reported in Asia but no population-specific nephrolithiasis models have been developed. We aimed to develop a simplified nephrolithiasis prediction model for the Korean population by using data from general medical practice.

Materials and Methods: This was a prospective, population-based cohort study in Korea. A total of 497,701 participants from the National Health Insurance Service–National Sample Cohort (NHIS-NSC) were enrolled from 2002 to 2010. A Cox proportional hazards model was used.

Results: During a median follow-up time of 8.5 years (range, 2.0–8.9 years) and among 497,701 participants, there were 15,783 cases (3.2%) of nephrolithiasis. The parsimonious model included age, sex, income grade, alcohol consumption, body mass index, total cholesterol, fasting blood glucose, and medical history of diseases. The Harrell's C-statistic was 0.806 (95% confidence interval [CI], 0.790–0.821) and 0.805 (95% CI, 0.782–0.827) in the derivation and validation cohorts, respectively.

Conclusions: The results of the present study imply that nephrolithiasis risk can be predicted by use of data from general medical practice and based on predictors that clinicians and individuals from the general population are likely to know. This model comprises modifiable risk factors and can be used to identify those at higher risk who can modify their lifestyle to lower their risk for nephrolithiasis. This study also offers an opportunity for external validation or updating of the model through the incorporation of other risk predictors in other settings.

Keywords: Epidemiology; Kidney calculi; Nephrolithiasis; Risk

INTRODUCTION

Nephrolithiasis or urolithiasis (kidney stone) is the presence of renal calculi caused by disruptions in the balance between the solubility and the precipitation of salts in the urinary tract and kidneys [1]. The incidence of nephrolithiasis peaks between 20 and 30 years of age [1], but varies by sex and race [2]. Men have a twofold risk of stone formation compared with women, with a peak incidence at 30 years of age, whereas women have a bimodal age distribution, with

peaks at 35 and 55 years [3].

Nephrolithiasis is a common disorder in developed countries and is considered to be a disease of affluence [4,5], with substantial direct and indirect costs among working-age adults [6] and with a reported lifetime prevalence of 10% to 12% in men and 5% to 6% in women [7,8]. Romero et al. [9] reported a 5.0% prevalence of nephrolithiasis in South Korea, whereas Kim et al. [10] reported an expected lifetime prevalence of 6.0% and 1.8% among Korean men and women, respectively. Recently, a worldwide increase in the occurrence of kidney stone disease has been reported [11], and an estimated 8% to 16% of the US population experience at least one symptomatic stone by the age of 70 years [12]. However, few epidemiologic studies have investigated urolithiasis in Asia [13].

The pathogenesis of kidney stone formation is complex and involves both metabolic and environmental risk factors [14]. Nephrolithiasis risk factors include male sex, age, race [9], high socioeconomic status [15], body mass index (BMI) [16], blood pressure levels [15,16], diabetes [16], gout [16], chronic kidney disease (CKD) [16], hyperparathyroidism [17], inflammatory bowel disease (IBD) [18], smoking [19], alcohol consumption [20], and metabolic syndrome [21]. Kidney stones (calculi) are mineral concretions in the renal calyces and pelvis, which may be free or attached to the renal papillae [22]. Stones that develop in the urinary tract (known as urolithiasis or nephrolithiasis) form when the urine becomes excessively supersaturated with respect to a mineral, leading to crystal formation, growth, aggregation, and retention within the kidneys [23]. In some circumstances, urine can also become supersaturated with certain relatively insoluble drugs or their metabolites, leading to crystallization in the renal collecting ducts or iatrogenic stones [24].

The prevalence and incidence of nephrolithiasis have increased in most Asian countries in previous decades [25]. Tae et al. [13] emphasized the urgent need for nephrolithiasis preventive efforts owing to the rapidly changing prevalence of nephrolithiasis in Korea. Although the prevalence and disease burden of nephrolithiasis have been increasing in Korea [26], to date no nephrolithiasis risk prediction equations have been developed for the Korean population. Some models have been developed for nephrolithiasis-related outcomes in some populations; however, many of the prediction models developed in a particular population may not perform well in other populations [27]. Here, we developed and validated prediction equations and a simplified risk score to estimate risk for nephrolithiasis in a Korean population using data from general medical practice and based on risk predictors that individuals from the general population are likely to know.

MATERIALS AND METHODS

We developed and validated risk prediction equations in accordance with guidelines and protocols recommended by TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) [28].

1. Study design and participants

This was a population-based prospective cohort study in Korea using National Health Insurance Service–National Sample Cohort (NHIS-NSC) data collected from January 1, 2002, to December 31, 2010. This cohort comprised members from different professions and demographic attributes, making it representative of the general Korean population. The database contains longitudinal anonymized patient records of all claims data, including diagnostic codes of diseases, treatment details, monthly insurance premiums, prescriptions, clinical laboratory results, physician visits, and demographic information. The diagnostic codes are based on the Korean Classification of Diseases, Sixth Revision (KCD-6), which is compatible with the International Classification of Diseases, Tenth Revision (ICD 10th Revision). The data are arranged on the basis of date of medical treatment and not date of claim [29]. A detailed description of the cohort profile has been published elsewhere [29]. This study was approved by the Seoul National University Institutional Review Board (certificate number: E1811/002-008). This study was based on anonymised health records with no personal identifiers. Therefore, there was no need for informed consent (no direct interaction with patients) and the study was exempted.

2. Data extraction and risk predictors

We extracted data on all risk factors and randomly allocated participants to the derivation and validation samples according to a split sample method using a ratio of 2:1. On the basis of literature reviews and established hypotheses, we extracted data including disease diagnoses, date of diagnosis, sex, age, insurance premium as a proxy for income grade (socioeconomic status), anthropometric measures, smoking status, alcohol consumption, physical activity, fasting blood glucose, total cholesterol, blood pressure, and premorbidities based on medical history (diabetes, hypertension, ulcerative colitis or IBD [Chrohn's disease], CKD, gout, hyperparathyroidism, and coronary artery disease). We imputed missing data by using covariate values measured at the nearest time points.

3. Assessment of covariates

BMI was categorized as <18.5 kg/m², ≥18.5 kg/m² to 25 kg/m², >25 kg/m² to 30 kg/m², and >30 kg/m²; smoking as never, former, and current smokers; and alcohol consumption as rarely (<2 times/month), moderate drinker (2 to 3 times/month), and heavy drinker (≥4 times/month). Physical activity was categorized based on frequency per week into low (none), moderately active (1 to 2 times/week), and very active or high (≥3 times/week). Socioeconomic status was categorized based on insurance premium on a scale of 100% to proxy income grade as low (<30%), medium (30% to 60%), and high (>60%). Baseline age was categorized as <25 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, and >54 years. Hypertension status was categorized as systolic blood pressure (SBP) <120 mm Hg and diastolic blood pressure (DBP) <80 mm Hg, SBP ≥120 to 139 mm Hg or DBP ≥80 to 89 mm Hg, SBP ≥140 to 159 mm Hg or DBP ≥90 to 99 mm Hg, and SBP ≥160 mm Hg or DBP ≥100 mm Hg or medication use due to hypertension. Fasting glucose was categorized as <100 mg/dL, 100 to 125 mg/dL, and ≥126 mg/dL or medication use due to diabetes. Total cholesterol was categorized as <200 mg/dL, 200 to 240 mg/dL, and >240 mg/dL. History of diagnosed medical conditions was based on the presence of ICD-10-CM records for diabetes (E10-14), hyperparathyroidism (E21.0-E21.5), hypertension (I10-I15), coronary artery disease or ischemic heart disease (I20-I25), CKD (N18.1-N18.9), gout (M10.0-M10.17), and IBD as selected related codes (K50-K52).

4. Outcome ascertainment and exclusion criteria

The outcome of interest was first diagnosis of nephrolithiasis. Nephrolithiasis (ICD-10-CM) diagnostic codes were extracted as codes N20 (calculus of kidney and ureter), N20.0 (calculus of kidney), N20.1 (calculus of ureter), and N20.2 (calculus of kidney with calculus of ureter). Participants who experienced at least one nephrolithiasis episode at baseline (before January 1, 2004) were excluded from the study. The earliest recorded date of nephrolithiasis diagnosis was the index date for the diagnosis. Participants were censored at the last recorded date, death, or study end date (December 31, 2010).

5. Statistical analyses

1) Descriptive statistics

We used Student's t-test for continuous variables and χ^2-test for categorical variables to examine the differences in baseline characteristics between participants in the derivation and validation cohorts stratified on the basis of nephrolithiasis.

2) Model derivation and construction of point-based risk scoring system

We used Cox proportional-hazard regression models to assess associations between risk predictors and nephrolithiasis and to derive prediction equations in the derivation sample. We defined time to event as the time from the first examination date to the date of first nephrolithiasis diagnosis, last recorded date, or date of death. Participants who were not diagnosed with nephrolithiasis or experienced death without nephrolithiasis were censored at date of death or end of study.

We initially conducted three analyses including univariate analysis, partially adjusted analysis, which adjusted for age and sex, and a fully adjusted analysis, which adjusted for age, sex, income grade, smoking status, physical activity, and alcohol consumption. We examined Cox proportional hazards assumptions and assessed the functional form of covariates and adopted clinically meaningful categories for nonlinear covariates. We used hierarchical cluster analysis and assessed estimated coefficients for predictors in the univariate analysis to select representative predictors for each cluster of correlated variables. The model was fitted and variables retained if they were significant at α=0.15 using a backward selection procedure. To construct a risk score, the estimated β coefficient for each variable was multiplied by 100 and rounded to the nearest integer.

3) Model validation and performance evaluation

Calibration is used to measure agreement between predicted probabilities and the actual outcomes. We used Hosmer–Lemeshow (H-L) type χ^2 (Nam and D'Agostino), which was calculated by dividing the data into deciles on the basis of the predicted probabilities from the model with the average predicted probabilities for each decile being compared with the actual risk probabilities of nephrolithiasis estimated by the Kaplan–Meier approach. We obtained the associated calibration plot. Discrimination is used to measure the model's ability to distinguish between non-events and events. The model discrimination performance was evaluated on the basis of Harrell's C-statistic, specificity, and sensitivity. We also determined the predictive accuracy (Brier score) and explained variation using Schemper–Henderson predictive measure (R^2). All analyses were conducted by using SAS version 9.4 (SAS Inc., Cary, NC, USA) and Python (version 3.7).

Table 1. Baseline characteristics of participants in the derivation and validation cohorts for nephrolithiasis (NPL)

Covariate	Derivation cohort (n=332,284)			Validation cohort (n=165,417)		
	Without NPL (n=321,733, 96.8%)	With NPL (n=10,551, 3.2%)	p-value	Without NPL (n=160,185, 96.8%)	With NPL (n=5,232, 3.2%)	p-value
Years of follow-up	8.5±1.0	5.3±2.0	<0.0001	8.5±1.0	5.3±2.0	<0.0001
Height (cm)	163.2±9.0	164.3±8.8	<0.0001	163.1±9.0	164.2±8.7	<0.0001
Weight (kg)	62.3±11.1	65.3±11.4	<0.0001	62.2±11.1	65.2±11.1	<0.0001
Body mass index (kg/m^2)	23.3±3.3	24.1±3.2	<0.0001	23.3±3.3	24.1±3.2	<0.0001
SBP (mm Hg)	122.8±17.2	124.5±16.7	<0.0001	122.8±17.2	124.6±16.6	<0.0001
DBP (mm Hg)	77.4±11.5	78.7±11.2	<0.0001	77.4±11.5	78.8±11.2	<0.0001
FBG (mg/dL)	93.9±28.9	95.1±28.8	<0.0001	94.0±28.9	95.0±28.8	0.0472
Total cholesterol (mg/dL)	192.1±38.8	196.7±38.6	<0.0001	192.1±38.7	196.5±37.8	<0.0001
Sex			<0.0001			<0.0001
Male	161,437 (50.2)	6,641 (62.9)		80,008 (49.9)	3,292 (62.9)	
Female	160,296 (49.8)	3,910 (37.1)		80,177 (50.1)	1,940 (37.1)	
Age (y)			<0.0001			<0.0001
<25	48,130 (15.0)	874 (8.3)		24,017 (15.0)	437 (8.2)	
25–34	66,014 (20.5)	2,041 (19.3)		32,839 (20.5)	977 (18.7)	
35–44	79,604 (24.8)	2,877 (27.3)		39,805 (24.8)	1,437 (27.5)	
45–54	60,932 (18.9)	2,553 (24.2)		30,056 (18.8)	1,213 (23.2)	
>54	67,053 (20.7)	2,206 (20.9)		33,468 (20.9)	1,168 (22.3)	
Income/insurance			<0.0001			<0.0001
Low (<30%)	47,741 (14.8)	1,302 (12.3)		23,613 (14.7)	652 (12.5)	
Medium (30%–60%)	115,835 (36.0)	3,620 (34.3)		57,734 (36.0)	1,811 (34.6)	
High (>60%)	158,157 (49.2)	5,629 (53.4)		78,838 (49.2)	2,769 (52.9)	
Physical activity/week			<0.0001			<0.0001
Low (none)	188,553 (58.6)	5,800 (55.0)		94,032 (58.7)	2,850 (54.5)	
Moderate (1–2 times/week)	113,704 (35.3)	4,036 (38.3)		56,329 (35.2)	2,010 (38.4)	
High (>3 times)	19,476 (6.1)	715 (6.8)		9,824 (6.1)	372 (7.1)	
Smoking status			<0.0001			<0.0001
Never	214,766 (66.8)	6,324 (59.9)		107,052 (66.8)	3,123 (59.7)	
Former smoker	14,058 (4.4)	567 (5.4)		7,024 (4.4)	307 (5.9)	
Current smoker	92,909 (28.9)	3,660 (34.7)		46,109 (28.8)	1,802 (34.4)	
Alcohol consumption			<0.0001			<0.0001
Rarely (<2 times)	163,826 (50.9)	4,910 (46.5)		81,759 (51.0)	2,483 (47.5)	
Moderate drinker (2–3 times)	130,770 (40.6)	4,651 (44.1)		64,901 (40.5)	2,251 (43.0)	
Heavy drinker (≥4 times)	27,137 (8.4)	990 (9.4)		13,525 (8.4)	498 (9.5)	
Body mass index (kg/m^2)			<0.0001			<0.0001
<18.5	20,057 (6.3)	349 (3.3)		10,051 (6.3)	158 (3.0)	
18.5–24.9	207,570 (64.5)	6,235 (59.1)		103,515 (64.6)	3,098 (59.2)	
25–29.9	85,608 (26.6)	3,554 (33.7)		42,382 (26.5)	1,800 (34.4)	
≥30	8,498 (2.6)	413 (3.9)		4,237 (2.6)	176 (3.4)	
FBG (mg/dL)			0.0771			0.7158
<100	201,610 (62.7)	6,520 (61.8)		100,270 (62.6)	3,250 (62.1)	
100–125	81,454 (25.3)	2,693 (25.5)		40,681 (25.4)	1,337 (25.6)	
≥126 or Tx.	38,669 (12.0)	1,338 (12.7)		19,234 (12.0)	645 (12.3)	
HTN			<0.0001			<0.0001
SBP <120 and DBP <80	121,113 (37.6)	3,418 (32.4)		60,373 (37.7)	1,696 (32.4)	
SBP 120–139 or DBP 80–89	164,789 (51.2)	5,788 (54.9)		81,875 (51.1)	2,875 (55.0)	
SBP 140–159 or DBP 90–99	31,953 (9.9)	1,206 (11.4)		15,938 (10.0)	584 (11.2)	
SBP ≥160 or DBP ≥100 or Rx.	3,878 (1.2)	139 (1.3)		1,999 (1.2)	77 (1.5)	

Table 1. Continued

Covariate	Derivation cohort (n=332,284)			Validation cohort (n=165,417)		
	Without NPL (n=321,733, 96.8%)	With NPL (n=10,551, 3.2%)	p-value	Without NPL (n=160,185, 96.8%)	With NPL (n=5,232, 3.2%)	p-value
Total cholesterol (mg/dL)			<0.0001			<0.0001
<200	190,372 (59.2)	5,763 (54.6)		94,775 (59.2)	2,866 (54.8)	
200–239	95,035 (29.5)	3,359 (31.8)		47,290 (29.5)	1,707 (32.6)	
>240	36,326 (11.3)	1,429 (13.5)		18,120 (11.3)	659 (12.6)	
Diagnosed IBD			0.0279			0.5226
No	320,492 (99.6)	10,496 (99.5)		159,570 (99.6)	5,209 (99.6)	
Yes	1,241 (0.4)	55 (0.5)		615 (0.4)	23 (0.4)	
Diagnosed CKD			0.5423			0.0107
No	321,396 (99.9)	10,542 (99.9)		160,030 (99.9)	5,221 (99.8)	
Yes	337 (0.1)	9 (0.1)		155 (0.1)	11 (0.2)	
Hyperparathyroidism			0.1916			0.5998
No	318,719 (99.1)	10,439 (98.9)		158,647 (99.0)	5,178 (99.0)	
Yes	3,014 (0.9)	112 (1.1)		1,538 (1.0)	54 (1.0)	
Diagnosed IHD			0.0032			0.0018
No	318,201 (98.9)	10,403 (98.6)		158,489 (98.9)	5,153 (98.5)	
Yes	3,532 (1.1)	148 (1.4)		1,696 (1.1)	79 (1.5)	
Diagnosed gout			<0.0001			0.0818
No	321,114 (99.8)	10,511 (99.6)		159,846 (99.8)	5,215 (99.7)	
Yes	619 (0.2)	40 (0.4)		339 (0.2)	17 (0.3)	

Values are presented as mean±standard deviation or number (%). The sum of the percentages does not equal 100% because of rounding.
SBP, systolic blood pressure; DBP, diastolic blood pressure; FBG, fasting blood glucose; Tx., medical utilisation due to diabetes; HTN, hypertension; Rx., medication utilisation due to hypertension; IBD, inflammatory bowel disease; CKD, chronic kidney disease; IHD, ischemic heart disease.

RESULTS

1. Baseline characteristics of the derivation and validation groups

The extracted data for 502,342 participants. We excluded 4,641 participants who had experienced at least one nephrolithiasis episode before January 1, 2004. During a median follow-up time of 8.5 years (range, 2.0–8.9 years) and among 497,701 participants, a total of 15,783 participants (3.2%) were diagnosed with nephrolithiasis. The total number of person-years of follow-up was 4,183,410 years. The mean (standard deviation) value for covariates and the distribution of the baseline characteristics stratified by nephrolithiasis in the derivation and validation cohorts are presented in Table 1. There were no discrepancies between the cohorts. However, there was a significant difference in baseline characteristics between those who developed nephrolithiasis and those who did not (p<0.05; Table 1).

2. Model derivation

Table 2 presents the estimated coefficients and hazard ratios (HRs) for each covariate in the univariate, partially adjusted, and fully adjusted analyses. On the basis of the univariate associations and after assessment of multicollinearity, a total of 14 covariates that were significantly associated with nephrolithiasis were assessed in the model derivation and retained if they were significant at α=0.15. The parsimonious model comprised age, sex, income grade, alcohol consumption, BMI, fasting blood glucose, IBD, total cholesterol, gout, and hyperparathyroidism. Table 3 presents the estimated coefficients and HRs for predictors in the parsimonious model. In the right-most column of the table are the points associated with the presence of a given level of a risk factor (with the reference level assigned zero points).

3. Model validation

The Harrell's C-statistics were 0.806 (95% confidence interval [CI], 0.790–0.821) and 0.805 (95% CI, 0.782–0.827) for the derivation and validation cohorts, respectively. A value of 0.50 represents no discrimination and 1.00 represents perfect discrimination. The model showed good calibration in the derivation and validation cohorts (H-L χ^2=8.6659, p=0.8879, and χ^2=8.5893, p=0.8351). The Brier scores were 0.0318 (95% CI, 0.0312–0.0324) and 0.0316 (95% CI, 0.0308–0.0325) in the derivation and validation cohorts, respectively. The Brier score ranges from 0.0 to 1.0 with lower values indicating

Table 2. Hazard ratios (HRs) (95% confidence interval [CI]) and β-coefficients for risk predictors in the univariate, partially adjusted, and fully adjusted analyses

Covariate	Unadjusted[a]			Partially adjusted[b]			Fully adjusted[c]		
	β (SE)	HR (95% CI)	p-value	β (SE)	HR (95% CI)	p-value	β (SE)	HR (95% CI)	p-value
Sex									
Male	0.557 (0.016)	1.75 (1.69–1.80)	<0.0001	0.560 (0.017)	1.75 (1.70–1.81)	<0.0001	0.565 (0.022)	1.76 (1.69–1.84)	<0.0001
Age (y)									
<25	Reference	Reference		Reference	Reference		Reference	Reference	
25–34	0.519 (0.033)	1.68 (1.58–1.79)	<0.0001	0.452 (0.033)	1.57 (1.47–1.68)	<0.0001	0.445 (0.033)	1.56 (1.46–1.67)	<0.0001
35–44	0.668 (0.032)	1.95 (1.83–2.07)	<0.0001	0.648 (0.032)	1.91 (1.80–2.03)	<0.0001	0.626 (0.032)	1.87 (1.76–1.99)	<0.0001
45–54	0.782 (0.032)	2.19 (2.05–2.33)	<0.0001	0.771 (0.032)	2.16 (2.03–2.30)	<0.0001	0.751 (0.032)	2.12 (1.99–2.26)	<0.0001
>54	0.591 (0.033)	1.81 (1.69–1.92)	<0.0001	0.597 (0.033)	1.82 (1.70–1.94)	<0.0001	0.579 (0.033)	1.78 (1.67–1.90)	<0.0001
Income/insurance premium									
Low (<30%)	Reference	Reference		Reference	Reference		Reference	Reference	
Medium (30%–60%)	0.130 (0.026)	1.14 (1.08–1.20)	<0.0001	0.072 (0.027)	1.07 (1.02–1.13)	0.0070	0.072 (0.027)	1.08 (1.02–1.13)	0.0066
High (>60%)	0.247 (0.025)	1.28 (1.22–1.35)	<0.0001	0.139 (0.025)	1.15 (1.09–1.21)	<0.0001	0.136 (0.025)	1.15 (1.09–1.20)	<0.0001
Physical activity									
High (>3 times)	Reference	Reference		Reference	Reference		Reference	Reference	
Moderate (1–2 times/week)	-0.043 (0.033)	0.96 (0.90–1.02)	0.1957	-0.017 (0.033)	0.98 (0.92–1.05)	0.6101	-0.015 (0.033)	0.99 (0.92–1.05)	0.6534
Low (none)	-0.195 (0.032)	0.82 (0.77–0.88)	<0.0001	-0.050 (0.032)	0.95 (0.89–1.01)	0.1256	-0.041 (0.032)	0.96 (0.90–1.02)	0.2034
Smoking status									
Never	Reference	Reference		Reference	Reference		Reference	Reference	
Former smoker	0.362 (0.035)	1.44 (1.34–1.54)	<0.0001	0.029 (0.037)	1.03 (0.96–1.11)	0.4295	0.033 (0.037)	1.03 (0.96–1.11)	0.3709
Current smoker	0.319 (0.017)	1.38 (1.33–1.42)	<0.0001	-0.010 (0.021)	0.99 (0.95–1.03)	0.6212	0.003 (0.021)	1.00 (0.96–1.05)	0.8987
Alcohol consumption									
Rarely (<2 times)	Reference	Reference		Reference	Reference		Reference	Reference	
Moderate drinker (2–3 times)	0.172 (0.017)	1.19 (1.15–1.23)	<0.0001	-0.040 (0.018)	0.96 (0.93–0.99)	0.0297	-0.041 (0.019)	0.96 (0.93–0.99)	0.0278
Heavy drinker (≥4 times)	0.217 (0.028)	1.24 (1.18–1.31)	<0.0001	-0.067 (0.030)	0.94 (0.88–0.99)	0.0265	-0.065 (0.030)	0.94 (0.88–0.99)	0.0318
Body mass index (kg/m²)									
<18.5	-0.563 (0.046)	0.57 (0.52–0.62)	<0.0001	-0.298 (0.046)	0.74 (0.68–0.81)	<0.0001	-0.296 (0.046)	0.74 (0.68–0.81)	<0.0001
18.5–24.9	Reference	Reference		Reference	Reference		Reference	Reference	
25–29.9	0.315 (0.017)	1.37 (1.32–1.42)	<0.0001	0.223 (0.017)	1.25 (1.21–1.29)	<0.0001	0.222 (0.017)	1.25 (1.21–1.29)	<0.0001
≥30	0.409 (0.043)	1.51 (1.39–1.64)	<0.0001	0.391 (0.043)	1.48 (1.36–1.61)	<0.0001	0.393 (0.043)	1.48 (1.36–1.61)	<0.0001
Fasting blood glucose (mg/dL)									
<100	Reference	Reference		Reference	Reference		Reference	Reference	
100–125	0.016 (0.019)	1.02 (0.98–1.05)	0.3875	-0.027 (0.019)	0.97 (0.94–1.01)	0.1470	-0.025 (0.019)	0.98 (0.94–1.01)	0.1890
≥126 or Tx.	0.050 (0.025)	1.05 (1.00–1.10)	0.0432	-0.030 (0.025)	0.97 (0.92–1.02)	0.2247	-0.025 (0.025)	0.97 (0.93–1.02)	0.3090

Table 2. Continued

Covariate	Unadjusted[a]			Partially adjusted[b]			Fully adjusted[c]		
	β (SE)	HR (95% CI)	p-value	β (SE)	HR (95% CI)	p-value	β (SE)	HR (95% CI)	p-value
Total cholesterol (mg/dL)									
<200	Reference	Reference		Reference	Reference		Reference	Reference	
200–239	0.152 (0.018)	1.17 (1.13–1.21)	<0.0001	0.080 (0.018)	1.08 (1.05–1.12)	<0.0001	0.077 (0.018)	1.08 (1.04–1.12)	<0.0001
>240	0.221 (0.024)	1.25 (1.19–1.31)	<0.0001	0.131 (0.025)	1.14 (1.09–1.20)	<0.0001	0.128 (0.025)	1.14 (1.08–1.19)	<0.0001
Hypertension (HTN)									
SBP <120 and DBP <80	Reference	Reference		Reference	Reference		Reference	Reference	
SBP 120–139 or DBP 80–89	0.219 (0.018)	1.25 (1.20–1.29)	<0.0001	0.065 (0.018)	1.07 (1.03–1.11)	0.0003	0.070 (0.018)	1.07 (1.04–1.11)	0.0001
SBP 140–159 or DBP 90–99	0.278 (0.028)	1.32 (1.25–1.39)	<0.0001	0.053 (0.028)	1.06 (1.00–1.12)	0.0609	0.063 (0.029)	1.07 (1.01–1.13)	0.0267
SBP ≥160 or DBP ≥100 or Rx.	0.266 (0.070)	1.31 (1.14–1.50)	0.0001	-0.009 (0.0701)	0.99 (0.86–1.14)	0.8986	0.006 (0.070)	1.01 (0.88–1.15)	0.9330
Premorbidities									
Ischemic heart disease	0.265 (0.067)	1.30 (1.14–1.49)	<0.0001	0.189 (0.067)	1.21 (1.06–1.38)	0.0050	0.185 (0.067)	1.20 (1.05–1.37)	0.0061
Inflammatory bowel disease	0.224 (0.114)	1.25 (1.00–1.56)	0.0480	0.213 (0.114)	1.24 (0.99–1.55)	0.0605	0.209 (0.114)	1.23 (0.99–1.54)	0.0657
Chronic kidney disease	0.246 (0.224)	1.28 (0.83–1.98)	0.2714	0.134 (0.224)	1.14 (0.74–1.77)	0.5492	0.125 (0.224)	1.13 (0.73–1.76)	0.5771
Hyperparathyroidism	0.082 (0.078)	1.09 (0.93–1.26)	0.2946	0.190 (0.078)	1.21 (1.04–1.41)	0.0149	0.184 (0.078)	1.20 (1.03–1.40)	0.0189
History of gout	0.567 (0.133)	1.76 (1.36–2.29)	<0.0001	0.299 (0.133)	1.35 (1.04–1.75)	0.0242	0.293 (0.133)	1.34 (1.03–1.74)	0.0274

Tx, medical utilisation due to diabetes; Rx., medication utilisation due to hypertension; SE, standard error; SBP, systolic blood pressure; DBP, diastolic blood pressure.
[a]:Univariate analysis. [b]:Partially adjusted for age, sex and each variable added. [c]:Fully adjusted accounting for age, sex, income grade, physical activity, smoking status, alcohol consumption and each of the other risk factors ([a]:fasting blood glucose/diabetes, total cholesterol, blood pressure/HTN, prior history of [ischemic heart disease, chronic kidney disease, hyperparathyroidism, inflammatory bowel disease and gout]).

higher prediction accuracy. The sensitivity and specificity in the validation cohort were 0.793 (95% CI, 0.740–0.875) and 0.510 (95% CI, 0.389–0.653), respectively. The equations explained 34% and 36% of the variation in time to diagnosis of nephrolithiasis in the derivation and validation cohorts, respectively. Table 4, Figs. 1, and 2 presents the model validation results. Subgroup analysis was conducted and model performance was comparable to the results from the main analyses.

4. Prediction equations

The individualized probability of diagnosis of nephrolithiasis within the years of follow-up (t=8) can be estimated by using the following equation:

P (Nephrolithiasis) $= 1 - S_o(t)^{\exp[f(x)]}$, where

$$f(x) = \sum_i \beta_i x_i$$

In the above equation, $S_o(t)$ denotes the baseline survival probability at time (t) for an individual with all covariates equivalent to zero (0), $(\beta_i's)$ denotes change in log hazard rate, and $(x_i's)$ denotes the values of the predictors.

5. Simplified risk score

The median score was 87, and the 25th and 75th percentiles were 61 and 118, respectively. The Youden's index suggested a risk score of ≥89 as the optimal cutoff to define high-risk individuals on the basis of the simplified risk score. This threshold identified 68.3% of individuals who developed nephrolithiasis, whereas application of the lower threshold at the 25th percentile identified 86.2% of the nephrolithiasis events.

6. Practical application of the nephrolithiasis risk score

Based on Table 5, the following example illustrates how nephrolithiasis risk can be estimated by using the simplified points system.

Case: a 52-year-old male with income grade of 30% to 60% who is a moderate drinker (2–3 times/month), is obese (BMI >30 kg/m²), has normal total cholesterol (<200 mg/dL), has normal fasting blood glucose (<100 mg/dL), and is without gout but has hyperparathyroidism and IBD.

For this case, $S_0(t)$ is nephrolithiasis-free average survival probability at the end of follow-up (t=8 years), estimated by Cox regression analysis. Based on the point system, the probability of nephrolithiasis can be estimated as follows:

P (Nephrolithiasis) $= 1 - (0.9545)^{\exp[(215/100)]} = 0.330$

Table 3. Hazard ratios (HRs) (95% confidence interval [CI]) and β-coefficients for risk predictors in the parsimonious model of nephrolithiasis and the risk point scoring system

Covariates	β coefficient (SE)	HR (95% CI)	p-value	Points[a]
Sex				
Female	Reference	Reference		0
Male	0.548 (0.023)	1.73 (1.65–1.81)	<0.0001	55
Age (y)				
<24	Reference	Reference		0
25–34	0.403 (0.041)	1.50 (1.38–1.62)	<0.0001	40
35–44	0.547 (0.039)	1.73 (1.60–1.87)	<0.0001	55
45–54	0.662 (0.040)	1.94 (1.79–2.10)	<0.0001	66
>54	0.466 (0.041)	1.59 (1.47–1.73)	<0.0001	47
Income grade/insurance premium				
Low (<30%)	Reference	Reference		0
Medium (30%–60%)	0.071 (0.033)	1.07 (1.01–1.14)	0.0283	7
High (>60%)	0.138 (0.031)	1.15 (1.08–1.22)	<0.0001	14
Alcohol consumption				
Rarely (<2 times/month)	Reference	Reference		0
Moderate drinker (2–3 times/month)	-0.029 (0.023)	0.97 (0.93–1.02)	0.1979	-3
Heavy drinker (>3 times/month)	-0.068 (0.037)	0.93 (0.87–1.01)	0.0664	-7
Body mass index (kg/m^2)				
<18.5	-0.261 (0.056)	0.77 (0.69–0.86)	<0.0001	-26
18.5–24.9	Reference	Reference		0
25–29.9	0.206 (0.022)	1.23 (1.18–1.28)	<0.0001	21
≥30	0.428 (0.051)	1.53 (1.39–1.70)	<0.0001	43
Total cholesterol (mg/dL)				
<200	Reference	Reference		0
200–239	0.043 (0.022)	1.04 (1.00–1.09)	0.0525	4
>240	0.106 (0.030)	1.11 (1.05–1.18)	0.0005	11
Fasting blood glucose (mg/dL)				
<100	Reference	Reference		0
100–125	-0.037 (0.023)	0.96 (0.92–1.01)	0.1056	-4
≥126 or Tx.	-0.042 (0.031)	0.97 (0.90–1.02)	0.1731	-4
Inflammatory bowel disease			0.0506	
No	Reference	Reference		0
Yes	0.264 (0.135)	1.30 (0.99–1.70)		26
Hyperparathyroidism			0.0253	
No	Reference	Reference		0
Yes	0.213 (0.095)	1.24 (1.03–1.49)		21
History of gout			0.0267	
No	Reference	Reference		0
Yes	0.352 (0.159)	1.42 (1.04–1.94)		35

SE, standard error; CI, confidence interval; Tx., medical utilisation due to diabetes.
[a]:The risk points were calculated by multiplying the estimated coefficients by 100 and rounding to the next integer.

DISCUSSION

We derived and validated prediction equations and a simplified risk score to estimate the risk for nephrolithiasis based on a combination of predictors that individuals are likely to know and that are routinely collected in general medical practice. The study was based on a large, repre-sentative Korean population from a validated nationwide database [29]. Risk prediction models derived from routinely collected health data are more readily applicable in clinical practice. This model is based on 10 risk predictors: age, sex, income grade, alcohol consumption, BMI, total cholesterol, fasting blood glucose, gout, IBD, and hyperparathyroidism. The model showed excellent calibration, a performance mea-

Table 4. Model validation and performance evaluation based on discrimination and calibration in derivation and validation cohorts

Performance evaluation statistic	Derivation cohort	Validation cohort
Brier score (95% CI)[a]	0.0318 (0.0312–0.0324)	0.0316 (0.0308–0.0325)
Schemper–Henderson (R^2)	0.34	0.36
Nam and D'Agostino test[b] (χ^2/p)	8.6659/0.8879	8.5893/0.8351
Harrell's C-statistic (95% CI)[c]	0.806 (0.790–0.821)	0.805 (0.782–0.827)
Sensitivity (95% CI)	0.789 (0.730–0.873)	0.793 (0.740–0.875)
Specificity (95% CI)	0.510 (0.389–0.652)	0.510 (0.389–0.653)

CI, confidence interval.
[a]:Measures both discrimination and calibration; range (0 to 1) and lower values indicate higher accuracy. [b]:A modification of Hosmer–Lemeshow test suited for survival data; measure of calibration that is specific to censored survival data (lower χ^2 and higher p-values) indicate better calibration. [c]:A measure of discrimination for which higher values indicate better discrimination.
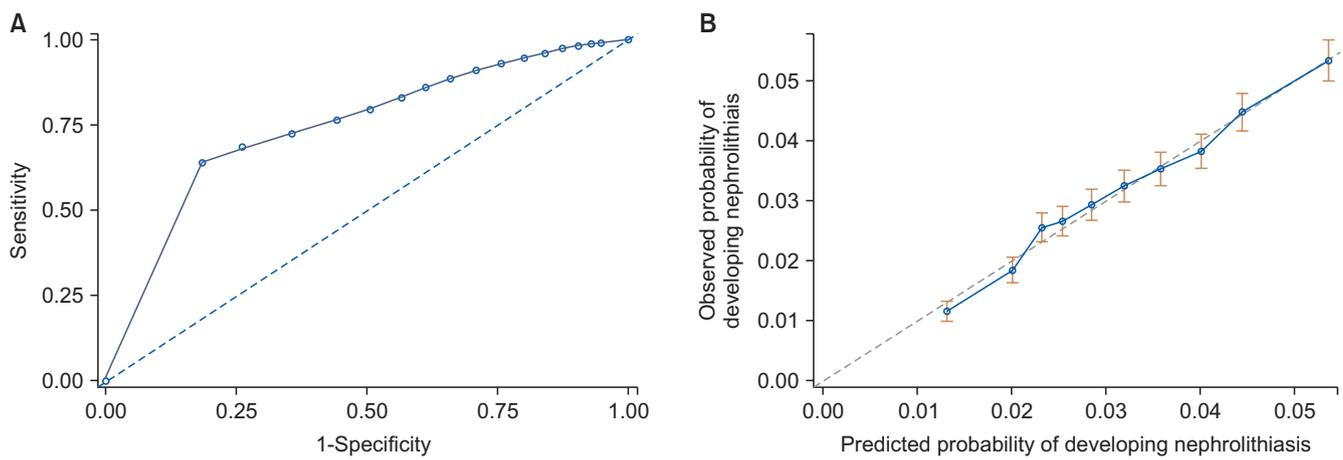
Fig. 1. Discrimination and calibration plots in the derivation cohort. (A) Discrimination. (B) Calibration.
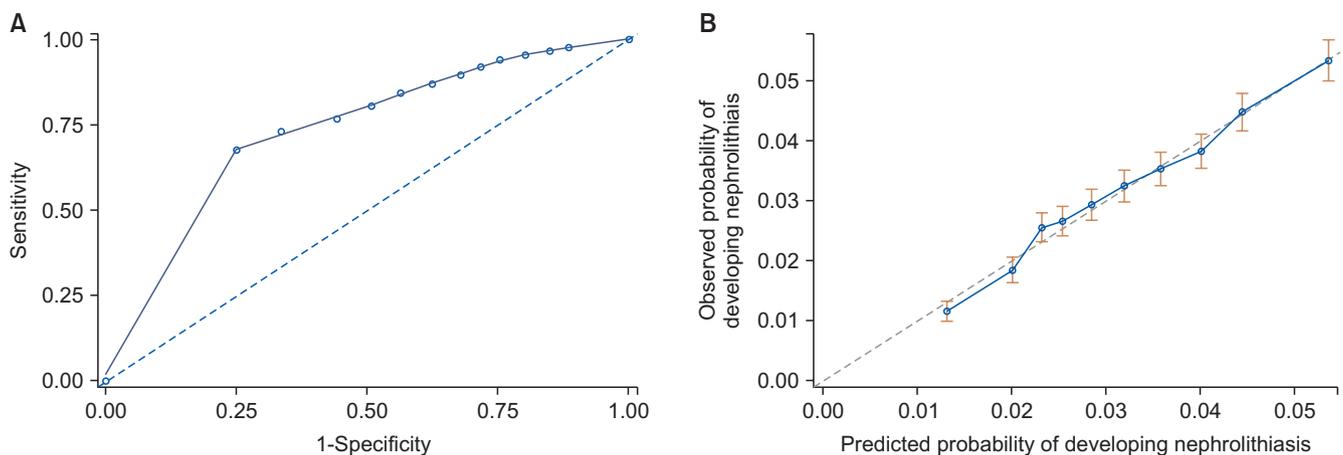
Fig. 2. Discrimination and calibration plots in the validation cohort. (A) Discrimination. (B) Calibration.

sure that is essential with regard to informing or making decisions in clinical practice. Model calibration primarily determines clinical utility together with the distribution of prediction around the optimum cutoff value and discrimination. The prediction equations also showed good discrimination, with a Harrell's C-statistic value of at least 0.805 in the validation cohort.

Knowledge of personalized risk for nephrolithiasis may motivate individuals to reduce their risks through appropriate interventions, thereby promoting population health and reducing personal and societal costs. This model can be used when a physician counsels an individual after a routine check-up to provide information about the nephrolithiasis risk profile and to give the exact probability of nephrolithia-

Table 5. Calculated score for a hypothetical example of a nephrolithiasis risk profile

Risk factor	Value (risk factor category)	Points
Sex	Male	55
Age (y)	45–54	66
Insurance premium	Medium (30%–60%)	7
Alcohol consumption (2–3 times/mo)	Moderate drinker	-3
Body mass index (kg/m^2)	≥ 30	43
Total cholesterol (mg/dL)	<200	0
Fasting blood glucose (mg/dL)	<100	0
Inflammatory bowel disease	Yes	26
History of hyperparathyroidism	Yes	21
History of gout	No	0
Total points		215
Estimate of risk		0.330

$S_o(t)=0.9545$.

sis. This may motivate lifestyle adjustments and promote adherence to treatment of premorbidities that are predictive of nephrolithiasis. Reducing risk factors associated with metabolic syndrome, proper therapy for individuals with metabolic syndrome, and management of hyperparathyroidism can greatly reduce nephrolithiasis risk. Lifestyle modification can reduce nephrolithiasis risk conferred through components of metabolic syndrome (prevention of complications from obesity, diabetes, and dyslipidemia). However, this prediction model was developed by using data from general medical practice and its performance may be higher in clinical practice settings than in the general population. Nevertheless, the ICD-10-CM codes in Korean health insurance claims data have good accuracy and correspondence with actual health status based on medical charts [30]; therefore, we believe the derived equations are applicable in medical practice.

There does not seem to exist a proper and confirmed method of preventing nephrolithiasis, but lifestyle and dietary adjustments may be helpful, and treatment of premorbidities may generally reduce nephrolithiasis risk. The equations will also improve self-awareness of general health status because the predictors in the model are also predictive of other health outcomes. The information could also be used by the government to reduce the burden of nephrolithiasis risk factors at the population level. Knowledge of the overall health status of a patient with respect to nephrolithiasis risk and expert knowledge from clinical practitioners will create a much clearer picture than either one alone. These variables in the model can easily be obtained in clinical practice, and the points system is simple to use.

Other studies have been done of clinical prediction to diagnose nephrolithiasis on the basis of symptoms, but our study focused on the risk of developing nephrolithiasis in apparently nephrolithiasis-free participants. Previous studies developed diagnostic models in patients suspected of having nephrolithiasis, in patients eligible for computed tomography, for symptomatic stone recurrence, and for determining the presence and type of renal stone. These studies were based on small samples and did not incorporate routinely collected medical data. Here, we have developed and validated prediction equations to estimate future risk of nephrolithiasis among apparently healthy individuals in a large cohort using routinely collected data. Furthermore, the existing models were developed in predominantly white populations and thus may not be appropriate and accurate in predicting nephrolithiasis risk among Asians. Racial differences in distribution and risk for nephrolithiasis have been reported in Asia. To our knowledge, the present study provides the first nephrolithiasis risk model for the Korean population. This model will improve individual decision-making, guide physicians in practice, and define groups at high risk for nephrolithiasis. This study also offers an opportunity for external validation of the model using data from other populations as well as the opportunity to update the model by incorporating other risk predictors in other settings, especially in this era of precision medicine.

This study has the strengths of representativeness, adequate sample size, long follow-up, and lack of selection, recall, and respondent biases. Data collected from health insurance claims in Korea have good accuracy [30], and the diagnosis of nephrolithiasis was made by clinicians and was not based on patient reports [13]. The Korea National Health Insurance Service (KNHIS) covers X-ray, intravenous pyelogram, and non-enhanced computed tomography imaging, and all physicians are required to provide evidence documenting nephrolithiasis [13]. Therefore, this study has considerable face validity.

The KNHIS database was not designed primarily for surveillance but supports other health care activities. Diagnostic suspicion and surveillance biases may occur because symptomatic or high-risk patients are more likely to undergo screening, and a differential intensity search for nephrolithiasis among patients with covariates associated with nephrolithiasis increases the probability of diagnosis. This systematic screening can result in detection of asymptomatic cases of nephrolithiasis in addition to symptomatic cases. Therefore, the incidence rate reported in this study may be higher than the number of symptomatic cases encountered in practice. Our study was also limited by known predictors that are not routinely recorded in medical practice, includ-

ing family history and dietary risk factors.

CONCLUSIONS

We have derived and validated prediction equations and a simplified nephrolithiasis risk score by using data representative of the entire Korean population. These prediction equations showed good performance and can be used in medical practice by health professionals to identify high-risk groups. Knowledge of individualized risk will motivate lifestyle adjustments and adherence to treatment of premorbidities, thereby reducing nephrolithiasis risk and improving general health.

CONFLICTS OF INTEREST

The authors have nothing to disclose.

AUTHORS' CONTRIBUTIONS

Research conception and design: David Mukasa. Data acquisition: Joohon Sung. Statistical analysis: David Mukasa. Data analysis and interpretation: David Mukasa and Joohon Sung. Drafting of the manuscript: David Mukasa. Critical revision of the manuscript: David Mukasa and Joohon Sung. Supervision: Joohon Sung. Approval of the final manuscript: all authors.

REFERENCES

1. Han H, Segal AM, Seifter JL, Dwyer JT. Nutritional Management of Kidney Stones (Nephrolithiasis). Clin Nutr Res 2015;4:137-52.

2. Curhan GC. Epidemiology of stone disease. Urol Clin North Am 2007;34:287-93.

3. Parmar MS. Kidney stones. BMJ 2004;328:1420-4.

4. Curhan GC, Willett WC, Rimm EB, Speizer FE, Stampfer MJ. Body size and risk of kidney stones. J Am Soc Nephrol 1998;9:1645-52.

5. Stapleton AE, Dziura J, Hooton TM, Cox ME, Yarova-Yarovaya Y, Chen S, et al. Recurrent urinary tract infection and urinary Escherichia coli in women ingesting cranberry juice daily: a randomized controlled trial. Mayo Clin Proc 2012;87:143-50.

6. Saigal CS, Joyce G, Timilsina AR; Urologic Diseases in America Project. Direct and indirect costs of nephrolithiasis in an employed population: opportunity for disease management? Kidney Int 2005;68:1808-14.

7. Stamatelou KK, Francis ME, Jones CA, Nyberg LM, Curhan GC. Time trends in reported prevalence of kidney stones in the United States: 1976-1994. Kidney Int 2003;63:1817-23.

8. Hiatt RA, Dales LG, Friedman GD, Hunkeler EM. Frequency of urolithiasis in a prepaid medical care program. Am J Epidemiol 1982;115:255-65.

9. Romero V, Akpinar H, Assimos DG. Kidney stones: a global picture of prevalence, incidence, and associated risk factors. Rev Urol 2010;12:e86-96.

10. Kim Hh, Jo MK, Kwak C, Park SK, Yoo KY, Kang D, et al. Prevalence and epidemiologic characteristics of urolithiasis in Seoul, Korea. Urology 2002;59:517-21.

11. Edvardsson VO, Indridason OS, Haraldsson G, Kjartansson O, Palsson R. Temporal trends in the incidence of kidney stone disease. Kidney Int 2013;83:146-52.

12. Scales CD Jr, Smith AC, Hanley JM, Saigal CS; Urologic Diseases in America Project. Prevalence of kidney stones in the United States. Eur Urol 2012;62:160-5.

13. Tae BS, Balpukov U, Cho SY, Jeong CW. Eleven-year cumulative incidence and estimated lifetime prevalence of urolithiasis in Korea: a national health insurance service-national sample cohort based study. J Korean Med Sci 2018;33:e13.

14. Sakhaee K, Maalouf NM, Sinnott B. Clinical review. Kidney stones 2012: pathogenesis, diagnosis, and management. J Clin Endocrinol Metab 2012;97:1847-60.

15. Arias Vega R, Pérula de Torres LA, Jiménez García C, Carrasco Valiente J, Requena Tapia MJ, Cano Castiñeira R, et al. Comorbidity and socio-demographic factors associated with renal lithiasis in persons aged 40 to 65: a cross-sectional study. Med Clin (Barc) 2017;149:383-90.

16. Ando R, Nagaya T, Suzuki S, Takahashi H, Kawai M, Okada A, et al. Kidney stone formation is positively associated with conventional risk factors for coronary heart disease in Japanese men. J Urol 2013;189:1340-6.

17. D'Angelo A, Calò L, Cantaro S, Giannini S. Calciotropic hormones and nephrolithiasis. Miner Electrolyte Metab 1997;23:269-72.

18. Cury DB, Moss AC, Schor N. Nephrolithiasis in patients with inflammatory bowel disease in the community. Int J Nephrol Renovasc Dis 2013;6:139-42.

19. Nakanishi N, Tatara K, Nakamura K, Suzuki K. Risk factors for the incidence of hyperuricaemia: a 6-year longitudinal study of middle-aged Japanese men. Int J Epidemiol 1999;28:888-93.

20. Wang X, Xu X, Wu J, Zhu Y, Lin Y, Zheng X, et al. Systematic review and meta-analysis of the effect of alcohol intake on the risk of urolithiasis including dose-response relationship. Urol Int 2015;94:194-204.

21. Jeong IG, Kang T, Bang JK, Park J, Kim W, Hwang SS, et al. Association between metabolic syndrome and the presence of kidney stones in a screened population. Am J Kidney Dis 2011;58:383-8.

22. Khan SR, Pearle MS, Robertson WG, Gambaro G, Canales BK, Doizi S, et al. Kidney stones. Nat Rev Dis Primers 2016;2:16008.

23. Finlayson B. Physicochemical aspects of urolithiasis. Kidney Int 1978;13:344-60.

24. Tattevin P, Revest M, Chapplain JM, Ratajczak-Enselme M, Arvieux C, Michelet C. Increased risk of renal stones in patients treated with atazanavir. Clin Infect Dis 2013;56:1186.

25. Liu Y, Chen Y, Liao B, Luo D, Wang K, Li H, et al. Epidemiology of urolithiasis in Asia. Asian J Urol 2018;5:205-14.

26. Choi SY, Lee SY, Chi BH, Kim JW, Kim TH, Chang IH. Urbanization may affect the incidence of urolithiasis in South Korea. Springerplus 2016;5:1891.

27. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.

28. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.

29. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol 2017;46:e15.

30. Park BJ, Sung J, Park K, Seo SW, Kim SH. Studying on diagnosis accuracy for health insurance claims data in Korea. Seoul: Seoul National University; 2003;17-29.