Original Article

Utility of Bayesian Logistic Regression Model in the Development of a Clinical Risk Score Index for Screening of Osteoporosis in Menopausal Women

Busaba Supawattanabodee PhD[1], Budsaba Wiriyasirivaj MD[2]

[1] Department of Basic Medical Science, Faculty of Medicine Vajira Hospital, Navamindradhiraj University, Bangkok, Thailand
[2] Department of Obstetrics and Gynecology, Faculty of Medicine Vajira Hospital, Navamindradhiraj University, Bangkok, Thailand

Objective: To evaluate the effectiveness of Bayesian logistic regression model in the development of a clinical risk score index for screening of osteoporosis in menopausal women.

Materials and Methods: Data of 973 menopausal women attending the menopause clinic at Faculty of Medicine Vajira Hospital between January 2002 and January 2008 were used as a derivation cohort. Age, body weight, menopausal duration, current estrogen use, previous low impact fracture, and lumbar and total hip bone mineral density [BMD] measurement by dual energy x-ray absorptiometry [DEXA] were used to develop a scoring system under 4 different scenarios. By using the Bayesian logistic regression model, the beta coefficients from the best fitting model of each scenario were transformed into simplified scoring algorithms in the derivation cohort. The diagnostic performance and their 95% confidence intervals [CI] from the best fitting model was determined.

Results: In the derivative cohort under scenario 4 (n = 300), the distribution pattern from all categories of 3 variables (age, body weight and estrogen use) stabilized distribution pattern within the fitted model. This model the narrowest 95% CI and smallest Monte Carlo [MC] errors when compared with scenarios 1 to 3 (n = 150, 200, 250). The scoring system was based on 3 variables of age (in year; <50 = 0, 50 to 59 = 0.5, 60 to 69 = 1, >70 = 1.5), body weight (in kilogram; >60 = 0, 50 to 59 = 1, <50 = 2), and current estrogen use (yes = 0, no = 2), showed a good discriminatory performance in identifying risk of osteoporosis in menopausal women. A score of 3.5 or greater yielded an area under the curve of 0.674 (95% CI = 0.604 to 0.744) with sensitivity of 70.6% (95% CI = 65.4 to 75.4), and specificity of 64.3% (95% CI = 58.8 to 69.7).

Conclusion: The Bayesian logistic regression model is an alternative and effective approach to identify postmenopausal women at risk for osteoporosis.

Keywords: Bayesian model, Logistic regression, Menopause, Osteoporosis screening

J Med Assoc Thai 2018; 101 (Suppl. 8): S193-S199
Website: http://www.jmatonline.com

Osteoporosis is one of the most silent and costly chronic diseases in post-menopausal women. This results from lack of estrogen which has direct effects on the bone density, increased fractures, resulting in serious consequences of reduced quality of life, morbidity, and even mortality. Due to the economic and clinical impacts of osteoporosis, early identification of osteoporosis followed by appropriate treatments plays an important role in reducing fracture risk.

Bone mineral density [BMD], measured by dual energy x-ray absorptiometry [DEXA], is a potential predictor of fracture risk and is used to diagnose osteoporosis based on WHO criteria[1]. However, the high cost of DEXA limits the use of BMD tests in some hospitals. In the past 10 years, several risk score indices

Correspondence to:
Supawattanabodee B. Department of Basic Medical Science, Faculty of Medicine Vajira Hospital, Navamindradhiraj University, Bangkok 10300, Thailand.
Phone: +66-81-9144875
E-mail: busabasupa@yahoo.com

How to cite this article: Supawattanabodee B, Wiriyasirivaj B. Utility of Bayesian Logistic Regression Model in the Development of a Clinical Risk Score Index for Screening of Osteoporosis in Menopausal Women. J Med Assoc Thai 2018;101;Suppl.8: S193-S199.

have been developed and are commonly used for osteoporosis screening. These scoring systems indices include potential predictors of osteoporosis, such as, ethnicity, family history of fracture, history of rheumatoid arthritis, age, weight, estrogen use, and menopausal duration. Some risk scoring indices are commonly used in European and Asian countries. The indices which are used in Thailand included the Osteoporosis Index of Risk [OSIRIS], Osteoporosis Risk Assessment Instrument [ORAI], Osteoporosis Self-Assessment Tool for Asians [OSTA], Khon Kaen Osteoporosis Study scoring system [KKOS] and Vajira Osteoporosis Risk Score Index [VORSI][2-9].

One of the designs for generating and testing the performance of scoring systems is split-cohort design. This split-cohort design was done to test the internal and external validity of the scoring system. The subjects in the split-cohort design are divided into different cohorts: the first cohort for tested or derivative model, and the second cohort for trained or validation model. The number of subjects in split-cohort designs varies according to the sample size in each study. Generally, it is either half-half, one-third[9], one and a half with a half cohort are respectively performed in derivative and test models[10]. Moreover, in the sample (cohort) designs, the statistical procedure for developing the scoring system is also considered when designing the study.

The standard method for developing osteoporosis scoring systems is to identify the associated factors with disease by univariate analysis. The significant factors from univariate analysis will be included in a multivariate analysis. Alternative means to deal with uncertainty is the Bayesian method. For over 20 years, Bayesian procedures have been implemented. Traditionally, logistic regression analysis is used[2-9], however this method does not take account of uncertainty in the model selected and uncertainty in its beta coefficients. When designing the study, all factors related to the risk of disease are collected. These factors create uncertainty in the study; all factors are generally defined and held to be constant during developing logistic regression model. This model leads to increased or decreased risk scores even whether patients have disease or not. in many clinical trials funded by either the pharmaceutical industry or the United States Food and Drug Administration (US FDA)[11], and even used in biomedical studies unrelated to clinical trials. Recently, the Bayesian regression model has become widely used in the design of screening tests, prediction and diagnostic testing. Burd et al[12] proposed Bayesian logistic regression model to generate prediction model for mortality based on injury ICD 9 coding. The model had better predictive performance than traditional Injury Severity Score [ICSS]. Muller et al[13] proposed a Bayesian approach to evaluate diagnostic test results in patients with Bovine tuberculosis in Africa. The key difference between traditional (frequentist) and Bayesian approaches is that the Bayesian approach includes the associated probability distribution. The distribution depends on all currently available information about the parameters; the posterior probabilities will be estimated after updating the model[14-16]. From the Bayesian idea, we present a Bayesian logistic regression method for handling uncertainty and building a scoring system in different scenarios with different numbers of subjects used in the derivative cohort. The aims of the study were to increase the internal validity of the model and reduce the number of subjects used in derivative model. Moreover, the traditional procedure to construct the model involved manual effort; but in this study the Markov Chain Monte Carlo [MCMC] was done instead.

Materials and Methods

Study sample

This descriptive study was approved by the institution's review board. We used a database of 973 menopausal women who attended menopause clinic at Faculty of Medicine, Vajira Hospital between January 2002 and January 2008. The scoring system was developed by using data of age, body weight, menopausal duration, current estrogen use, previous low impact fracture, and lumbar and total hip bone mineral density [BMD] measurement by dual energy X-ray absorptiometry [DEXA].

Definition and notations

In this study, all of the factors are categorized as the following:

Age was categorized into four groups: (1) <50 years (reference group); (2) 50 to 59 years; (3) 60 to 69 years; (4) ≥70 years. The beta coefficients of these variables are b.age 1, b.age 2, b.age 3, and b.age 4.

Body weight was categorized into three groups: (1) ≥60 kgs (reference group); (2) 50-59 kgs; (3) <50 kgs. The beta coefficients of these variables are b.wt 0, b.wt 1, and b.wt 2.

Current estrogen use was categorized into two groups: (1) used (reference group); (2) never used. The beta coefficients of these variables are b.er 0, b.er.

Previous low impact fracture was categorized into two groups: (1) never had fracture (reference group); (2) had fracture. The beta coefficients of these variables are b.fr 0, b.fr.

Simulation process

The derivative model was done by using Bayesian logistic regression model on four different groups of subjects in each scenario as follows:

Scenario 1: n = 150
Scenario 2: n = 200
Scenario 3: n = 250
Scenario 4: n = 300

Under simulation process, Markov Chain Monte Carlo [MCMC] was used. Each scenario was repeated 20,000 times with WinBUGS Package Version 1.4. The prior distributions for all factors (age, body weight, menopause duration, current estrogen use, previous low impact fracture) are normal distribution.

Define mean of prior distributions = 0 and higher variance =1,000.

The performance of the model was evaluated by comparing the number of subjects required in each scenario, the MC error, and 95% confidence interval (CI) of mean of beta coefficients. The beta coefficients of all factors from the best fitting model were transformed into simplified scoring algorithms, the performance of these scoring systems was tested in the derivative cohort. Finally, diagnostic performance was tested by considering sensitivity, specificity, positive predictive value, negative predictive value, with 95% CI, and area under the curve [AUC] of receiver operating characteristic curve analysis.

Results

Part 1: simulation results

In the simulation study of each scenario, age, body weight and previous estrogen use were significantly associated with osteoporosis. The mean, standard deviation, 95% CI, MC error of beta coefficients from significant variables under four scenarios are shown in Table 1.

Under scenario 4 (n = 300), the distribution pattern from all categories of three variables (age, body weight and estrogen use) stabilized distribution pattern with fitted model as shown in Figure 1. This model produces the narrowest range in 95% CIs and smallest MC errors when compared with scenario 1 to 3 respectively (n = 150, 200, 250). The mean of beta coefficients, 95% CIs and MC errors are shown as the following;

(1) Age group 50-59 years or b.age2 = 0.316 (95% CI = -1.110, 2.024) with MC error = 0.033, age group 60-69 years or b.age 3 = 1.179 (95% CI = -0.237 , 2.900) with MC error = 0.034 , and age group >= 70 years or b.age4 = 1.588 (95% CI = 0.204, 3.252) with MC error = 0.034.

(2) Body weight group 50-59 kgs or b.wt1 = 0.740 (95% CI = -0.034, 1.550) with MC error = 0.007, body weight group < 50 kgs or b.wt2= 1.826 (95% CI = 0.875, 2.825) with MC error = 0.010.

(3) Never used estrogen or b.er = 2.037 (95% CI = 0.836, 3.627) with MC error = 0.037.

Part 2: scoring system and diagnostic performance

The beta coefficients from scenario 4 were transformed into a simplified scoring system as shown in the last column of Table 2. The cut-off point score ≥3.5 is better than score ≥4 for screening osteoporosis. This cut-off point yielded higher sensitivity = 70.6% (95% CI = 65.4 to 75.4), specificity = 64.3% (95% CI = 58.8 to 69.7), positive predictive value [PPV] = 28.8% (95% CI = 23.7 to 33.9), negative predictive value [NPV] = 91.4% (95% CI = 88.3 to 94.6), and higher AUC = 0.674 (95% CI = 0.604 to 0.744). Although cut-off point score ≥3 has higher sensitivity than score ≥3.5 (78.4% and 70.6%), this cut-off point has a lower specificity (51.8 and 64.3%) and smaller area under the curve (0.651 and 0.674).

Discussion

In previous studies, the split-sample or split-cohort design was used for screening, diagnosis or prediction designs[19,10,12,17]. The number of subjects in split-cohort design varies according to both derivative (training) and validation (testing) cohort. First, in a study of predicted mortality based on injury ICD 9 coding, the sample was divided into training and testing data. There were 447,442 medical records in training data, while 312,592 medical records were assigned into training data[12]. Second, in a study that screened for postmenopausal osteoporosis using Vajira Osteoporosis Risk Score Index [VORSI], 386 subjects were assigned into derivative cohort while 587 subjects were assigned into validation cohort[9]. Third, in a study predicting risk of HCC in asymptomatic individuals who were seropostive for anti-HCV antibody[10], there were 975 and 572 anti-HCV seropositives in derivative and validation model. After scoring system was performed by using beta coefficient from the logistic regression model, these scoring systems would be re-trained again in the validation or testing cohort. The advantage from

Table 1. Comparing the performance of the model from different scenarios

Scenario	Parameters	Mean	SD	MC error	2.5%	Median	97.5%
1 (n = 150)	alpha	-5.607	1.489	0.067	-8.884	-5.447	-3.121
	b.age 2	-0.057	1.074	0.035	-2.049	-0.099	2.217
	b.age 3	-0.131	1.123	0.036	-2.182	-0.171	2.229
	b.age 4	1.160	1.055	0.035	-0.753	1.107	3.413
	b.er	3.472	1.222	0.052	1.534	3.309	6.284
	b.wt 1	0.472	0.594	0.009	-0.665	0.4643	1.665
	b.wt 2	2.141	0.736	0.012	0.737	2.129	3.621
2 (n = 200)	alpha	-5.700	1.254	0.058	-8.370	-5.611	-3.474
	b.age 2	0.541	0.968	0.036	-1.220	0.492	2.628
	b.age 3	1.235	0.978	0.037	-0.5370	1.176	3.325
	b.age 4	2.035	0.940	0.036	0.333	1.972	4.063
	b.er	2.371	0.857	0.031	0.978	2.285	4.282
	b.wt 1	0.842	0.511	0.0089	-0.138	0.830	1.882
	b.wt 2	2.196	0.605	0.012	1.035	2.189	3.412
3 (n = 250)	alpha	-6.074	1.245	0.060	-8.827	-5.983	-3.909
	b.age 2	0.628	0.925	0.035	-1.037	0.571	2.606
	b.age 3	1.203	0.930	0.037	-0.470	1.146	3.237
	b.age 4	2.082	0.907	0.037	0.456	2.020	4.076
	b.er	2.633	0.842	0.033	1.222	2.554	4.534
	b.wt 1	1.043	0.471	0.009	0.168	1.033	2.013
	b.wt 2	2.173	0.571	0.011	1.086	2.160	3.323
4 (n = 300)	alpha	-5.106	1.232	0.067	-7.328	-4.993	-3.259
	b.age 2	0.316	0.824	0.033	-1.110	0.263	2.024
	b.age 3	1.179	0.832	0.034	-0.237	1.125	2.900
	b.age 4	1.588	0.809	0.034	0.204	1.533	3.252
	b.er	2.037	0.830	0.037	0.836	1.951	3.627
	b.wt 1	0.740	0.405	0.007	-0.034	0.733	1.550
	b.wt 2	1.826	0.497	0.010	0.875	1.817	2.825

split-sample design is the models are trained and tested in the same time with same sample with less time use. However, when the sample size is not large enough with multiple risk factors, the ratio of split-sample design is very controversial. Our study used only one cohort with various scenarios for the number of subjects under simulation process to derive the model.

The results of our study reveal that the best scenario (scenario 4) requires smaller sample size in derivative cohort (n = 300) when compared with VORSI study (n = 386)[9]. The results of beta coefficient from our study are similar to the VORSI study for two variables (body weight and estrogen use), while data on age shows a small difference. The beta coefficients for body weight and estrogen use groups between our study and VORSI are: (1) weight group 50 to 59 kgs = 0.740 and 0.711 (95% CI = -0.034, 1.550 and 0.033, 1.389); (2) weight group <50 kgs = 1.826 and 1.638 (95% CI = 0.875, 2.825 and 1.183, 2.296); (3) never used estrogen = 2.037 and 1.400 (95% CI = 0.836, 3.627 and 0.641, 2.159); (4) age group 50 to 59 years = 0.316 and 0.857 (95% CI = -1.110, 2.024 and -0.080, 1.794); (5) age group 60 to 69 years = 1.179 and 1.707 (95% CI = -0.237, 2.900 and 0.704, 2.709); (6) ≥70 years = 1.588 and 1.949 (95% CI = 0.204, 3.252 and -0.463, 3.761). The main reason for the difference in age group beta coefficients between our study and VORSI was that the sample size in age ≥70 was too small (n = 4 and 2); the Bayesian approach takes account of the uncertainty of this effect. The scoring system from our study is quite similar to the VORSI study, which transformed beta coefficients to integers for all variables, while our study transformed to integers and one decimal in some variables (e.g. age group). For that reason, the different cut-off points between our study and VORSI were selected (score ≥3.5 and score ≥4) to identify risk of osteoporosis. The results of diagnostic performance from our study are similar when compared to VORSI: sensitivity of = 70.6% and 71.9% (95% CI = 65.4 to 75.4 and 59.3 to 82.0); specificity = 64.3% and 62.1% (95% CI = 58.8 to 69.7

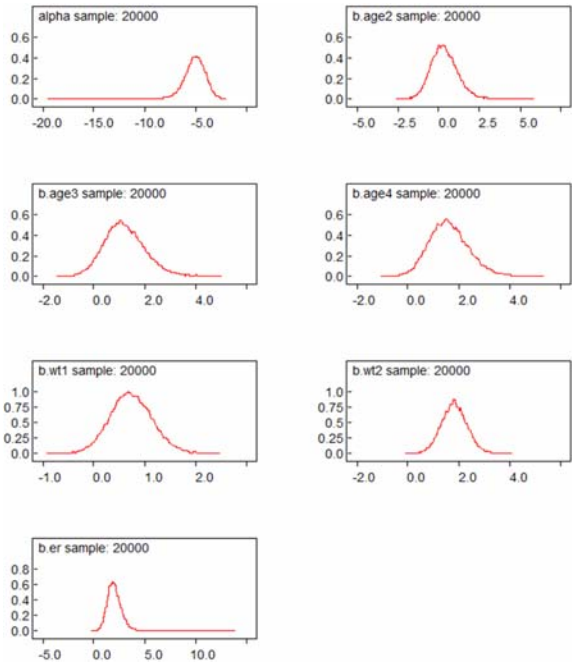

Figure 1. The distribution of the best fitting model (scenario 4: n = 300).

Table 2. Scoring system from the best-fitting model (scenario 4: n = 300)

Variables	Beta coefficients	Score
Age (years)		
<50 (n = 49)	Reference group	0
50 to 59 (n = 179)	0.316	0.5
60 to 69 (n = 68)	1.179	1
≥70 (n = 4)	1.588	1.5
Body weight (kg)		
≤60 (n = 166)	Reference group	0
50 to 59 (n = 140)	0.740	1
<50 (n = 44)	1.826	2
Current estrogen use		
Ever (n = 73)	Reference group	0
Never (n = 227)	2.037	2

and 57.0 to 67.1); PPV = 28.8% and 27.4% (95% CI = 23.7 to 33.9 and 24.4 to 30.0); NPV = 91.4% and 91.7% (95% CI = 88.3 to 94.6 and 90.0 to 95.1); and higher AUC = 0.674 and 0.750 (95% CI = 0.604 to 0.744 and 0.690 to 0.810). In summary, our proposed model (Bayesian logistic regression) is a useful statistical method to screen for osteoporosis in menopausal women.

Table 3. Sensitivity analysis for different cut off points in derivative model (n = 300)

Cut-off points	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC (95% CI)
Score ≥3	78.4% (73.8 to 83.1)	51.8% (46.1 to 57.5)	25.0% (20.1 to 29.9)	92.1% (89.1 to 95.2)	0.651 (0.586 to 0.716)
Score ≥3.5	70.6% (65.4 to 75.4)	64.3% (58.8 to 69.7)	28.8% (23.7 to 33.9)	91.4% (88.3 to 94.6)	0.674 (0.604 to 0.744)
Score ≥4	47.1% (44.4 to 52.7)	85.1% (81.1 to 89.2)	39.3% (33.8 to 44.9)	88.7% (85.1 to 92.3)	0.660 (0.588 to 0.734)

Remark: prevalence = 17.0%, 95% CI = (12.7 to 21.2)
AUC = area under the curve, CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value

Conclusion

Bayesian logistic regression model is an optional approach for screening, diagnosis, and prediction; and effectively deals with the uncertainty of data. This procedure can be applied in practice when the number of subjects is not large (around <300) while multiple risk factors have been concerned. The advantage of the Bayesian logistic procedure is that it requires fewer subjects to derive the model and offers similar or better diagnostic performance in identifying risk when compared with traditional logistic regression.

What is already known on this topic?

Bone mineral density [BMD], measured by dual energy x-ray absorptiometry [DEXA], is used to diagnose osteoporosis based on WHO criteria[1] but not widely implemented due to lack of budget in some hospital. In the past, several risk score indices have been developed and commonly used for screening osteoporosis with split cohort design. This conventional design together with logistic regression model has been used to establish and validate scoring system until now.

What this study adds?

The Bayesian logistic regression model was proposed to deal with uncertainty problem and low resource setting instead of the traditional logistic regression model. This Bayesian logistic regression model was allowed to test and validate in the different setting which was flexible than traditional logistic regression model. This study was also demonstrated the advantage of model in the different scenarios of sample size. The results also revealed that the Bayesian regression model was the alternative statistical method using with low resource setting especially when the sample size and budget was limited.

Potential conflicts of interest

The authors declare no conflict of interest.

References

1. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. World Health Organ Tech Rep Ser 1994;843:1-129.
2. Ribot C, Pouilles JM, Bonneu M, Tremollieres F. Assessment of the risk of post-menopausal osteoporosis using clinical factors. Clin Endocrinol (Oxf) 1992;36:225-8.
3. National Osteoporosis Foundation. Physician's guide to prevention and treatment of osteoporosis. Belle Mead, NJ: Excerpta Medica; 1998.
4. Lydick E, Cook K, Turpin J, Melton M, Stine R, Byrnes C. Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. Am J Manag Care 1998;4:37-48.
5. Black DM, Palermo L, Abbott T, Johnell O. SOSFSURF: A simple, useful risk factor system can identify the large majority of women with osteoporosis. Bone 1998;23:S605.
6. Cadarette SM, Jaglal SB, Kreiger N, McIsaac WJ, Darlington GA, Tu JV. Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry. CMAJ 2000;162:1289-94.
7. Weinstein L, Ullery B. Identification of at-risk women for osteoporosis screening. Am J Obstet Gynecol 2000;183:547-9.
8. Sedrine WB, Chevallier T, Zegels B, Kvasz A, Micheletti MC, Gelas B, et al. Development and assessment of the Osteoporosis Index of Risk (OSIRIS) to facilitate selection of women for bone densitometry. Gynecol Endocrinol 2002;16:245-50.
9. Wiriyasirivaj B, Supawattanabodee B. Development and internal validation of the Vajira Osteoporosis Risk Score Index (VORSI). Vajira Med J 2011;55:95-109.Ashby D. Bayesian statistics in medicine: a 25 year review. Stat Med 2006;25:3589-631.
10. Lee MH, Lu SN, Yuan Y, Yang HI, Jen CL, You SL, et al. Development and validation of a clinical scoring system for predicting risk of HCC in asymptomatic individuals seropositive for anti-HCV antibodies. PLoS One 2014;9:e94760.
11. Ashby D. Bayesian statistics in medicine: a 25 year review. Stat Med 2006;25:3589-631
12. Burd RS, Ouyang M, Madigan D. Bayesian logistic injury severity score: a method for predicting mortality using international classification of disease-9 codes. Acad Emerg Med 2008;15:466-75.
13. Muller B, Vounatsou P, Ngandolo BN, Diguimbaye-Djaibe C, Schiller I, Marg-Haufe B, et al. Bayesian receiver operating characteristic estimation of multiple tests for diagnosis of bovine tuberculosis in Chadian cattle. PLoS One 2009;4:e8215.
14. Dmitrienko A, Wang MD. Bayesian predictive approach to interim monitoring in clinical trials. Stat Med 2006;25:2178-95.

15. Bolstad WM. Introduction to Bayesian statistics. 2nd ed. New Jersey: John Wiley and Sons; 2007.
16. Emerson SS, Kitellson MJ, Gillen LD. Tutorial in biostatistics Frequentist evaluation group sequential trial designs. Stat Med 2007;26: 5047-80.
17. Rassi A Jr, Rassi A, Little WC, Xavier SS, Rassi SG, Rassi AG, et al. Development and validation of a risk score for predicting death in Chagas' heart disease. N Engl J Med 2006;355:799-808.