

Correlates and Predictors of NT-proBNP in Life Insurance Applicants

Steven J. Rigatti, MD, DBIM, DABFM; Robert Stout, PhD

Objectives.—To document the various laboratory and demographic/historical correlates of NT-proBNP levels in applicants for life insurance, and to explore the accuracy of a prediction model based on those variables.

Method.—NT-proBNP blood test results were obtained from 1.34 million insurance applicants between the age of 50 and 85 years, beginning in 2003. Exploratory data analysis was carried out to document correlations with other laboratory variables, sex, age, and the presence of relevant diseases. Further, predictive models were used to quantify the proportion of the variance of NT-proBNP, which can be explained by a combination of these other, easier to determine variables.

Results.—NT-proBNP shows the expected, negative correlation with estimated glomerular filtration rate (eGFR) is markedly higher in those with a history of heart disease and is somewhat higher in those with a history of hypertension. A strong, unexpected, negative correlation between NT-proBNP and albumin was discovered. Of the variables evaluated, a multivariate adaptive regression spline (MARS) model automated selection procedure selected 7 variables (age, sex, albumin, eGFR, BMI, systolic blood pressure, cholesterol, and history of heart disease). Variable importance evaluation determined that age, albumin and eGFR were the 3 most important continuous variables in the prediction of NT-proBNP levels. An ordinary least squares (OLS) model using these same variables achieved a R-squared of 24.7%.

Conclusion.—Expected ranges of NT-proBNP may vary substantially depending on the value of other variables in the prediction equation. Albumin is significantly negatively correlated with NT-proBNP levels. The reasons for this are unclear.

Address: Rigatti Risk Analytics, LLC, 157 Uconn Ave, Glastonbury, CT 06033, steven.rigatti@crlcorp.com

Correspondent: Steven J. Rigatti, MD.

Key words: NT-proBNP, BNP, linear models, modeling.

Author Affiliations: Rigatti – Founder, Rigatti Risk Analytics, LLC, Consultant Medical Director, Clinical Reference Laboratories, Lenexa KS. Stout – Chief Scientific Officer/Laboratory Director, Clinical Reference Laboratories, Lenexa KS.

Received: October 17, 2022

Accepted: March 3, 2023

Amino terminal pro-B-type natriuretic protein (NT-proBNP) is the non-active amino-terminal fragment of proBNP, a pro-hormone produced by the heart in response to left ventricular strain.^{1,2} The active, carboxy-terminus fragment, BNP, was first recognized as a clinically useful analyte in the differentiation of

heart failure from pneumonia in patients presenting to the emergency room with shortness of breath and varying degrees of hypoxia, leukocytosis and radiographic abnormalities on chest X-ray.³

Since then, it has been recognized that NT-proBNP has a longer half-life and is more

Table 1. Baseline Characteristics

	Female (n = 262,839)	Male (n = 480,788)	Overall (n = 743,627)
Age (y)			
Median [Q1, Q3]	63.0 [58.0, 70.0]	62.0 [57.0, 68.0]	63.0 [57.0, 69.0]
Status			
Alive	262,019 (99.7%)	478,664 (99.6%)	740,683 (99.6%)
Dead	820 (0.3%)	2,124 (0.4%)	2,944 (0.4%)
History of Heart Disease			
No	253,640 (96.5%)	445,791 (92.7%)	699,431 (94.1%)
NS	2,643 (1.0%)	6,231 (1.3%)	8,874 (1.2%)
Yes	6,556 (2.5%)	28,766 (6.0%)	35,322 (4.7%)
History of Diabetes			
No	237,365 (90.3%)	424,348 (88.3%)	661,713 (89.0%)
NS	2,433 (0.9%)	5,582 (1.2%)	8,015 (1.1%)
Yes	230,41 (8.8%)	50,858 (10.6%)	73,899 (9.9%)
History of Cancer			
No	163,108 (62.1%)	284,455 (59.2%)	447,563 (60.2%)
NS	87,706 (33.4%)	175,014 (36.4%)	262,720 (35.3%)
Yes	12,025 (4.6%)	213,19 (4.4%)	33,344 (4.5%)
History of Hypertension			
No	169,859 (64.6%)	29,6429 (61.7%)	466,288 (62.7%)
NS	2,821 (1.1%)	6,263 (1.3%)	9,084 (1.2%)
Yes	90,159 (34.3%)	178,096 (37.0%)	268,255 (36.1%)
Albumin (g/dl)			
Median [Q1, Q3]	4.40 [4.20, 4.60]	4.50 [4.30, 4.60]	4.40 [4.30, 4.60]
Creatinine (mg/dl)			
Median [Q1, Q3]	0.800 [0.710, 0.910]	1.02 [0.920, 1.14]	0.960 [0.820, 1.09]
eGFR (ml/min/1.73 m²)			
Median [Q1, Q3]	90.5 [85.4, 94.9]	98.5 [85.5, 109]	93.7 [85.5, 103]
Cholesterol (mg/dl)			
Median [Q1, Q3]	202 [178, 228]	184 [159, 210]	190 [165, 217]
BMI (kg/m²)			
Median [Q1, Q3]	25.7 [22.6, 29.6]	27.4 [25.0, 30.3]	26.9 [24.2, 30.1]
Systolic BP (mmHg)			
Median [Q1, Q3]	120 [112, 127]	123 [117, 130]	121 [115, 129]
Diastolic BP (mmHg)			
Median [Q1, Q3]	74.0 [69.0, 79.0]	77.0 [71.0, 81.0]	76.0 [71.0, 80.0]
NT-proBNP (pg/ml)			
Median [Q1, Q3]	65.7 [36.3, 118]	41.0 [22.0, 79.0]	49.0 [26.0, 94.0]

NS = Not stated

stable in serum than BNP, making it more suitable for testing. In this publication and others, NT-proBNP has been recognized as a strong predictor of mortality,^{4,5} not only in the insurance setting, but also in various clinical contexts, most notably congestive heart failure,⁶ valvular heart disease⁷ and stable coronary disease.^{8,9} Recently, NT-proBNP has attracted

attention as a potential marker of heart failure suitable for clinical monitoring, though evidence of the superiority of this approach has been lacking.¹⁰ Complicating the interpretation of elevated NT-proBNP levels is the fact that normal values vary by age, sex,¹¹ renal function,¹² and BMI.¹³ These other factors may affect NT-proBNP levels to

Table 2. Correlations

	r (95%CI)	
Age	0.409 (0.407, 0.411)	
Albumin	-0.238 (-0.240, -0.236)	
Creatinine	0.042 (0.039, 0.044)	
eGFR	-0.269 (-0.271, -0.267)	
log(Cholesterol)	-0.079 (-0.082, -0.077)	
log(BMI)	-0.059 (-0.062, -0.057)	
Systolic BP	0.068 (0.065, 0.070)	
Diastolic BP	-0.058 (-0.060, -0.055)	
	Mean Difference in log(NT-proBNP)(95% CI)	NT-proBNP %change
Heart Disease*	0.797(0.786,0.807)	122%
Cancer*	0.273 (0.262, 0.284)	31%
Hypertension*	0.239 (0.234, 0.243)	27%
Diabetes*	0.118 (0.110, 0.125)	13%

*Mean difference = $\log(\text{NT-proBNP} \mid \text{history} = \text{"Yes"}) - \log(\text{NT-proBNP} \mid \text{history} = \text{"No"})$

**Mean difference = $\log(\text{NT-proBNP} \mid \text{status} = \text{"Dead"}) - \log(\text{NT-proBNP} \mid \text{status} = \text{"Alive"})$

% increase in NT-proBNP when going from Yes (disease present) to No (disease absent)

an extent that it becomes difficult to interpret borderline elevations. Also, with so many valid correlates, one may wonder if the NT-proBNP level might be accurately predicted from other, more easily obtained variables.

The purpose of this article is to explore and quantify these correlates of NT-proBNP, with the goal of describing how these various quantities interact to account for some fraction of the total NT-proBNP level. Also, an online application is produced that predicts the mean level and higher quantiles of NT-proBNP when supplied with the predictor variables.

METHODS

NT-proBNP test results were collected from subjects applying for life insurance. Analysis was restricted to those aged 50-85, since very few NT-proBNP levels are available outside that range. For a subset of these, height, weight, and blood pressure were measured at the time of blood collection. Blood pressures were entered as averages of up to 3 repeated measurements. All applicants filled out a questionnaire with simple yes/no questions, asking about a personal history of heart

disease, hypertension, diabetes and cancer. Those not answering lab slip questions were categorized as "NS" (not stated). Collected blood was tested for NT-proBNP, as well as other substances depending on the individual life insurers' testing protocol. Analysis was restricted to those who had non-missing values for all selected numeric variables, which included age, BMI, systolic and diastolic blood pressure, serum creatinine, estimated glomerular filtration rate by the CKD-Epi equation,¹⁴ cholesterol, and albumin.

Correlations with NT-proBNP were studied in univariate and multivariate analyses. Because of a highly skewed distribution, NT-proBNP was log-transformed before inclusion in correlation studies or linear models. Correlation was explored with Pearson r tests for continuous variables and t-tests for categorical variables (non-responders were censored for the purpose of the t-tests). The variables of eGFR and systolic blood pressure were noted to be correlated with age ($r = -0.41$ for eGFR; $r = 0.16$ for systolic blood pressure). Therefore, linear models were used to remove the contribution of age and sex from these variables, producing age-corrected eGFR and systolic blood pressure. Variable selection for

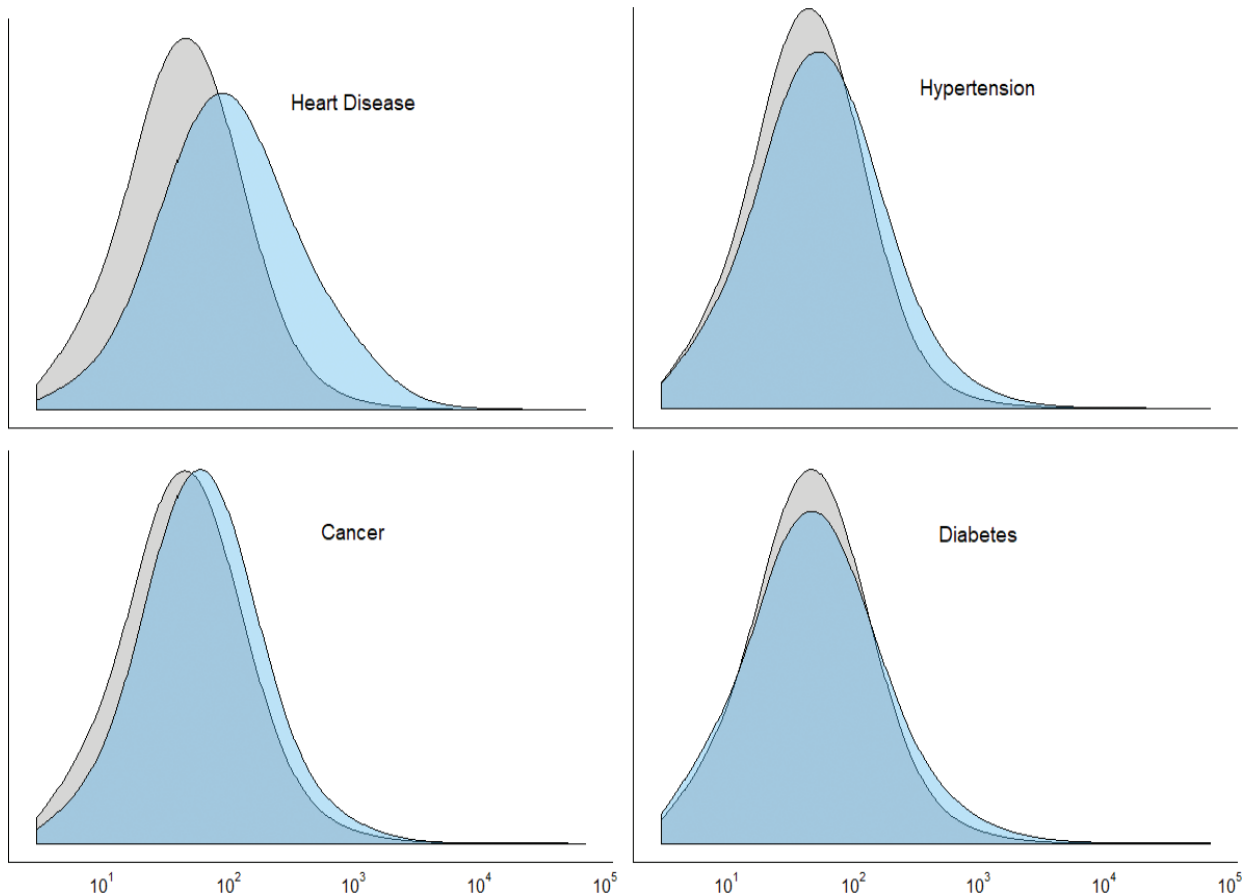


Figure 1. NT-proBNP by disease status.

inclusion in the final model was performed using multivariate adaptive regression splines (MARS) with a grid search for the degree of interaction (up to 3) and the number of included terms (up to 27). Variable importance was computed using change in the residual sum of squares with inclusion/exclusion of the variable, scaled to 100 for the most important variable. The final model utilized ordinary least squares (OLS) linear regression and was restricted to only those terms that showed non-zero importance in the MARS model. To enable non-linear effects, restricted cubic splines were utilized with 4 default knots in the OLS analysis.

Analyses were carried out using R version 4.1, and the following packages: tidyverse, rms, nephron, broom, skimr, matrixStats, table 1, caret, vip, and recipe.

RESULTS

The baseline characteristics of the study group are displayed in Table 1. The average age in the study was 63 years and was similar between men and women. The prevalence of heart disease, diabetes and hypertension were higher among men (5.8%, 10.6% and 36.9%, respectively), than among women (2.5%, 8.9%, and 34.3%). Cancer prevalence was marginally higher among women (4.6% vs 4.3%). The non-response rate was higher for cancer (35.3%) than the others (near 1%) because the cancer question was only added to the lab slip recently. Median levels of albumin, creatinine, eGFR, BMI, and blood pressure were all higher in men. Cholesterol was slightly higher in women as was the median level of NT-proBNP.

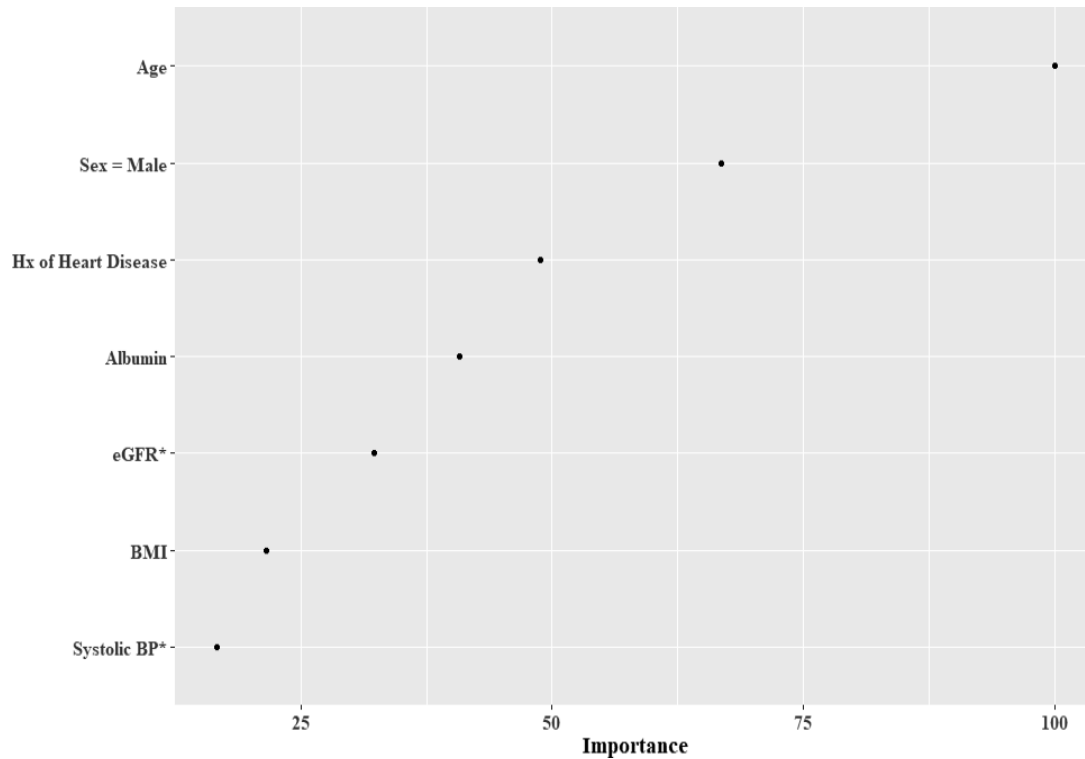


Figure 2. Variable importance (by residual sum of squares) for MARS model.

Table 3. Regression Model Results

Term	Estimate	Std. Error	t-statistic	<i>p</i>
(Intercept)	8.29	0.0858	96.7	<10 ⁻⁵
Age	0.03	0.0008	41.4	<10 ⁻⁵
Age'	0.03	0.0023	11.0	<10 ⁻⁵
Age''	-0.04	0.0076	-4.9	<10 ⁻⁵
Sex (M vs. F)	-0.35	0.0029	-118.9	<10 ⁻⁵
Heart Disease (NS vs. N)	0.09	0.0128	7.4	<10 ⁻⁵
Heart Disease (Y vs. N)	0.64	0.0056	114.2	<10 ⁻⁵
eGFR	-0.01	0.0002	-55.8	<10 ⁻⁵
eGFR'	0.01	0.0006	20.8	<10 ⁻⁵
eGFR''	-0.05	0.0067	-7.9	<10 ⁻⁵
Albumin	-0.88	0.0141	-62.5	<10 ⁻⁵
Albumin'	0.79	0.0513	15.4	<10 ⁻⁵
Albumin''	-1.92	0.1866	-10.3	<10 ⁻⁵
BMI	-0.05	0.0011	-46.6	<10 ⁻⁵
BMI'	0.09	0.0041	22.8	<10 ⁻⁵
BMI''	-0.20	0.0130	-15.6	<10 ⁻⁵
Systolic BP	0.00	0.0004	-6.0	<10 ⁻⁵
Systolic BP'	0.01	0.0011	10.2	<10 ⁻⁵
Systolic BP''	-0.02	0.0048	-4.5	<10 ⁻⁵

'First spline, '' Second Spline. M = Male, F = Female, NS = not stated, Y = yes, N = No, eGFR = estimated glomerular filtration rate, BMI = body mass index, BP = blood pressure

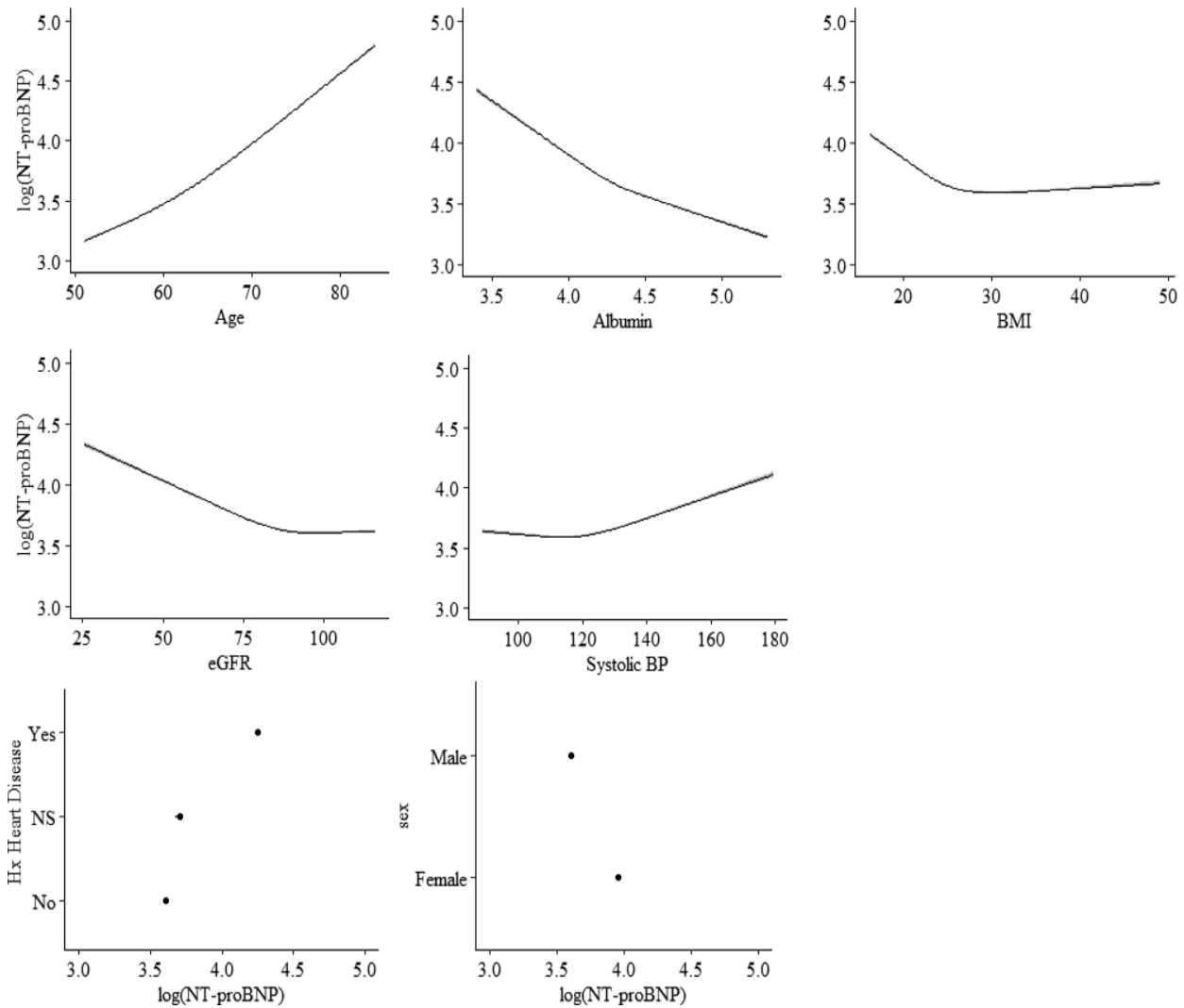


Figure 3. Model of main effects and spline terms.

Essentially all numeric variables had statistically significant correlations with NT-proBNP (Table 2), but this is not unexpected due to the large sample size. The strongest correlations were found with age (0.416), albumin (-0.25), and eGFR (-0.26). For all categorical variables, statistically significant differences in the mean log (NT-proBNP) level were found between “Yes” and “No” answers. The largest difference was found for heart disease (0.839 log units). In Table 2, the mean differences in log units are translated to a mean % change as one goes from “No” to “Yes”. Thus, the 131% change for heart disease history means that the level of NT-proBNP in those with a history of heart

disease is, on average, more than double the level in those without. Figure 1 demonstrates the distribution of NT-proBNP values by disease status. For all conditions, positive status results in a thicker rightward “tail” to the distribution, though the effect is much stronger for heart disease than the others.

Before modeling, the data were split into training (80%) and test (20%) sets, to provide more generalizable error estimates. The MARS model automated process selected 7 variables (age, sex, albumin, eGFR, BMI, systolic blood pressure, cholesterol, and history of heart disease), as well as 2 additional interaction terms (age-sex and age-eGFR). When evaluated using residual sum of squares for

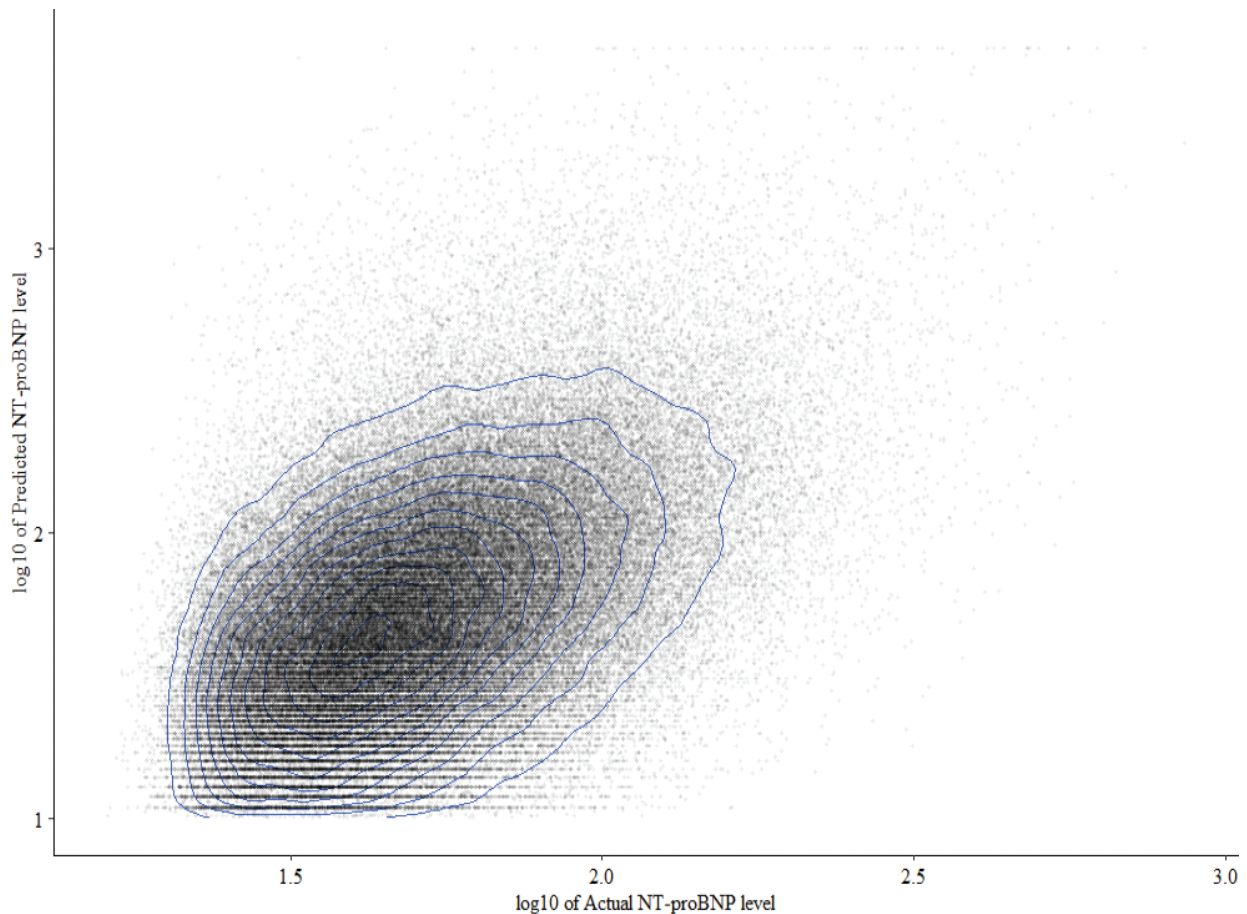


Figure 4. Predicted vs actual values from test data.

variable importance, the interactions terms had zero importance and were removed from consideration (Figure 2).

The variables selected from the MARS model were then utilized in an ordinary least squares (OLS) model where all continuous variables were entered using restricted cubic splines with 4 default knots. In this final model, all terms including main effects and spline terms were significant at the 99.9% confidence level (Table 3). A plot of partial effects (Figure 3) demonstrates a very strong effect of age across the included range. The effect of albumin is also quite strong and mostly log-linear. BMI appears to be associated with elevated NT-proBNP only when the BMI is below about 27 kg/m². Similarly, there is some evidence of a threshold for eGFR near

90 mg/ml/1.73m², and for systolic BP near 130 mmHg.

The final model has an R-squared of 25.7%, which means that the predictors, when combined in the manner described, can account for about a fourth of the variation in NT-proBNP levels in this population. The root mean square error on test data was 188.4, meaning that, on average, the prediction equation is in error by approximately 188 pg/ml. However, it is more informative to examine the distribution of prediction errors. This is done in Figure 4, which plots predicted vs actual values from the test data, and in Figure 5, which is a histogram of the individual prediction errors. We can see from these plots, that the OLS model is more likely to have highly negative errors (predict a value

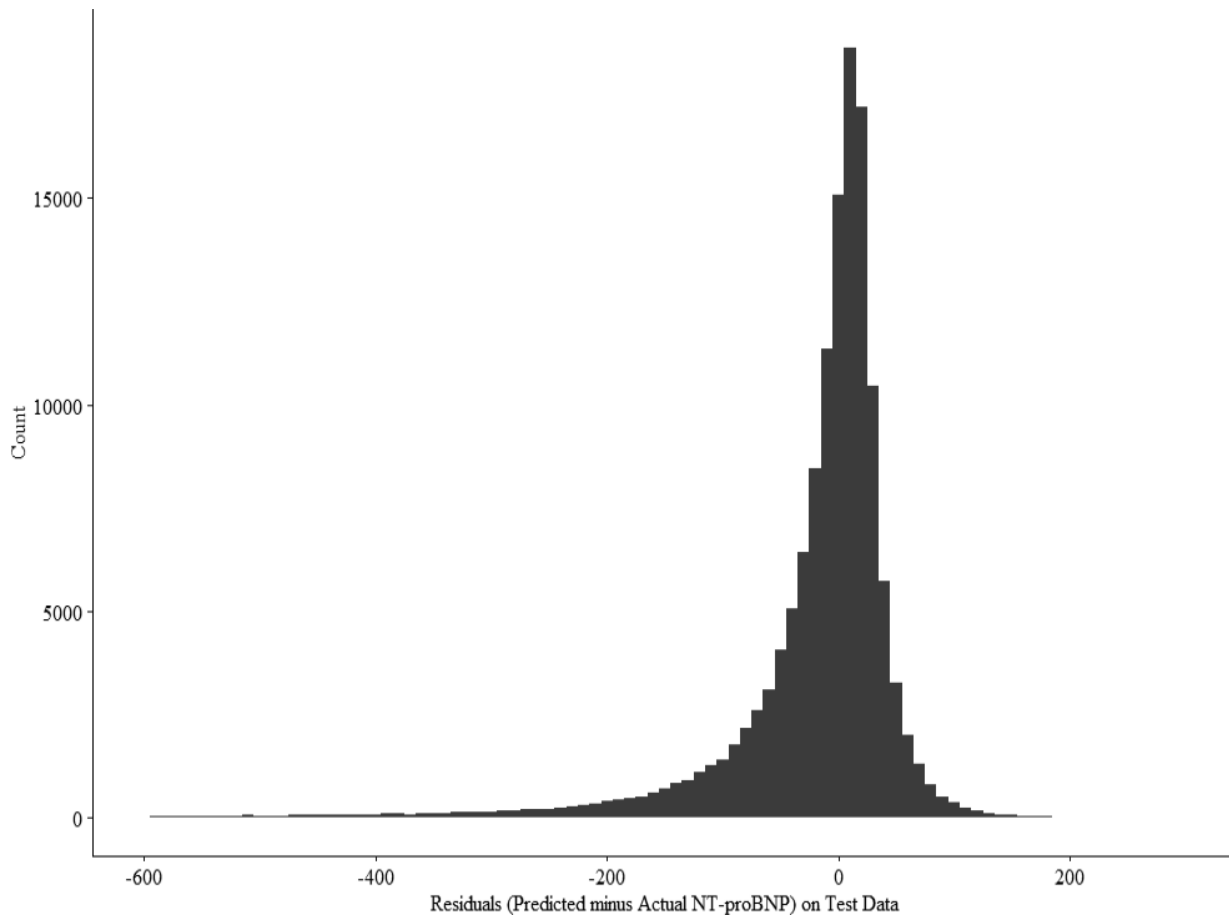


Figure 5. A histogram of the individual prediction errors.

that is far too low) than a highly positive error (predicting a value which is far too high).

DISCUSSION

The results presented here confirm the findings of many other studies that have shown a dependence of NT-proBNP on age, renal function, heart disease history, sex, and BMI. A surprising finding is the strong dependence on the albumin level. This finding has not been previously reported in medical literature, and any rationale can only be speculative. What we see is a fairly strong dependence on albumin levels, even well within the normal range, so it would not appear to be a signal of those illnesses which cause markedly low albumin levels (cancer, cachexia, etc). It may be that serum albumin

may interfere with the NT-proBNP assay, but this has not been described. It is also conceivable that there is some sort of interaction or binding to albumin within the serum. A recent study found a correlation between high NT-proBNP levels and the risk of cancer.¹⁵ Although the authors did control for various laboratory values, albumin was not one of them. There is a significant likelihood that the finding of association is confounded by the correlation between high NT-proBNP and low levels of albumin, which are known to be associated with elevated cancer death rates.

The overall performance of the model suggests that NT-proBNP is not simply an expression of other physiological phenomena as measured by laboratory testing. Rather, much as expected, NT-proBNP brings unique value, which cannot be accurately predicted

from other values. However, the degree to which it can be predicted is, perhaps, surprising since about a quarter of its variation is due to other measured factors. Likely, though, prediction is insufficiently accurate to justify a testing strategy based on the output of the OLS model. For instance, among those who were predicted to have NT-proBNP levels below 50 pg/ml, 15% had actual levels over 200 pg/ml.

In the insurance context, this study can help modify expectations for the NT-proBNP level in applicants who have various combinations of the other predictors, and to account for the new finding of a negative correlation with albumin level. For instance, an 80-year-old man with no history of heart disease, a BMI of 28 kg/m², a systolic BP of 130 mmHg, a creatinine of 1.1 mg/dl, and an albumin of 4.5 g/dl would have a predicted NT-proBNP level of 110 pg/ml. Increasing the creatinine to 1.7 mg/dl results in a predicted NT-proBNP of 152 pg/ml. A similar NT-proBNP of 154 pg/ml would be predicted if the creatinine remained at 1.1 mg/dl, but the albumin was lowered to 4 g/dl.

To help medical directors and underwriters with the various influencers of the NT-proBNP level, an interactive application has been created at <https://srigatti.shinyapps.io/BNPpredict/>. This application can take any combination of inputs from the predictors discussed in this paper and returns a prediction for the NT-proBNP level. It also uses prediction errors to predict higher quantiles of NT-proBNP distribution. This can help determine which levels may be abnormally high based on the other characteristics of the individual.

REFERENCES

- Daniels LB, Maisel AS. Natriuretic peptides. *J Am Coll Cardiol*. 2007;50:2357-2368.
- Hall C. Essential biochemistry and physiology of (NT-pro)BNP. *Eur J Heart Fail*. 2004;6:257-260.
- Maisel AS, Clopton P, Krishnaswamy P, et al. Impact of age, race and sex on the ability of B-type natriuretic peptide to aid in the emergency diagnosis of heart failure: results from the Breathing Not Properly (BNP) multinational study. *Am Heart J*. 2004;147:1078-1084.
- Clark M, Kaufman V, Fulks M, et al. NT-proBNP as a Predictor of All-Cause Mortality in a Population of Insurance Applicants. *J Insur Med*. 2014;44:7-16.
- Simsek MA, Degertekin M, Turer Cabbar A, et al. NT-proBNP levels and mortality in a general population-based cohort from Turkey: a long-term follow-up study. *Biomark Med*. 2018;12:1073-1081.
- Kang SH, Park JJ, Choi DJ, et al. Prognostic value of NT-proBNP in heart failure with preserved versus reduced EF. *Heart*. 2015;101:1881-1888.
- Gomez Peres M, Ble M, Cladellas M, et al. Combined use of tissue Doppler imaging and natriuretic peptides as prognostic marker in asymptomatic aortic stenosis. *Int J Cardiol*. 2017;228:890-894.
- Ruwald MH, Goetze JP, Bech J, et al. NT-proBNP independently predicts long-term mortality in patients admitted for coronary angiography. *Angiology*. 2014;65:31-36.
- Niccoli G, Conte M, Marchitti S, et al. NT-proANP and NT-proBNP circulating levels as predictors of cardiovascular outcome following coronary stent implantation. *Cardiovasc Revasc Med*. 2016;17:162-168.
- Khan MS, Siddiqi T J, Usman MS, et al. Does natriuretic peptide monitoring improve outcomes in heart failure patients? A systematic review and meta-analysis. *Int J Cardiol*. 2018;263:80-87.
- Redfield MM, Rodeheffer RJ, Jacobsen SJ, et al. Plasma brain natriuretic peptide concentration: impact of age and gender. *J Am Coll Cardiol*. 2002;40:976-982.
- McCullough PA, Duc P, Omland T, et al. B-type natriuretic peptide and renal function in the diagnosis of heart failure: an analysis from the Breathing Not Properly Multinational Study. *Am J Kidney Dis*. 2003;41:571-579.
- Suthahar N, Meijers WC, Ho JE, et al. Sex-specific associations of obesity and N-terminal pro-B-type natriuretic peptide levels in the general population. *Eur J Heart Fail*. 2018;20:1205-1214.
- Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604-612.
- Tuñón J, Higuera J, Tarín N, et al. N-Terminal Pro-Brain Natriuretic Peptide Is Associated with a Future Diagnosis of Cancer in Patients with Coronary Artery Disease. *PLoS ONE*. 2015;10: e0126741.