



# Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R

Jiaqi Liang, Guoshu Bi, Cheng Zhan

Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China

Correspondence to: Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai, China.

Email: czhan10@fudan.edu.cn.

*Provenance and Peer Review:* This article was commissioned by the editorial office, *Annals of Translational Medicine*. The article did not undergo external peer review.

*Comment on:* Zhou ZR, Wang WW, Li Y, *et al.* In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann Transl Med* 2019;7:796.

Submitted Mar 17, 2020. Accepted for publication Apr 28, 2020.

doi: 10.21037/atm-2020-57

View this article at: <http://dx.doi.org/10.21037/atm-2020-57>

It is essential in a standard linear regression analysis that dependent variables are continuous. However, using the standard linear regression for the analysis of a double-level or multi-level outcome can lead to unsatisfactory results because the validity of this regression model relies on the variability of the outcome being the same for all values of predictors, which is contrary to the nature of double-level or multi-level outcomes (1). Therefore, when the dependent variable consists of several categories, a maximum likelihood estimator, such as multinomial logit or probit, should be used instead of the ordinary least square estimator (2). Logistic regression can be used to describe the relationship between an independent variable(s) (either continuous or not) and a dichotomous or multi-categorical dependent variable as a supplementary variable to the standard linear regression.

Zhou *et al.* (3) elaborated a series of reliable methodologies using the R software to construct clinical prediction models with detailed steps and operable code examples, according to different types of clinical data and categories of variables. They summarized the process of the construction of practical clinical prediction models (nomograms), including data screening, primary model training, and internal and external validations, which was an extraordinary work and a practical reference in the field of statistics (4-6). Based on this study, a simpler and more accurate prediction model was introduced as an extension by Bi *et al.* (7), which was designed for extracting polynomial equations and

calculating the points of each variable together with survival probabilities.

The general objective of logistic regression models is to predict outcomes using variables based on certain existing data, which have been applied in medical research for various diseases (8). In the study by Zhou and his colleagues (3), the authors converted multi-categorical outcomes into dichotomous ones and introduced a dichotomous logistic regression using R codes. However, multi-categorical outcomes can be directly applied in multinomial or ordinal logistic regression analyses in the R software, although the results might be difficult to be interpreted with more complicated steps. This study aimed to display the methods and processes used to apply multi-categorical variables in logistic regression models in the R software environment.

The sample data was made up of patients registered in the SEER database in 2015 with diagnoses of lung adenocarcinomas. Patients with unclear race, primary site(s) of their tumors, differentiation grade of their tumors, tumor stage (AJCC, 6<sup>th</sup> edition), or cause of their death were excluded. Finally, 6,483 patients met the inclusion and exclusion criteria, of which 3,000 were randomly selected for the following analysis. Age, sex, race, primary site of the tumor, cell differentiation grade, AJCC stage of the tumor, and history of chemotherapy, were chosen as independent variables. The ages of included patients, ranging from 20 to 100 years, were divided into eight groups, labeled as "AgeGroup". Also, several

subgroup variables were defined for each of the other six variables: “Sex” (male and female), “Race” (white, black, and others), “PrimarySite” (main, upper lobe, middle lobe, lower lobe, and overlapped), “Grade” (I–IV), “Stage” (I–IV), and “Chemotherapy” (yes and no or unknown). Survival status labeled as “Causespecificdeathclassification” was defined as the dependent variable consisting of three categories: alive, dead due to lung cancer, and dead due to other causes. Survival months were extracted for constructing an ordinal outcome. Using the dplyr package in R, the survival months were converted into “SurvivalStatus” based on the tertiles of survival months. Patient outcomes were defined as “0 = no response” for those whose survival months are in the first third tertile; “1 = partial response” and “2 = complete response” were defined as outcomes for patients whose survival months are in the middle and last third tertiles, respectively.

Multinomial logistic regressions can be applied for multi-categorical outcomes, whereas ordinal variables should be preferentially analyzed using an ordinal logistic regression model. Besides, if the ordinal model does not meet the parallel regression assumption, the multinomial one will still be an alternative (9). A multinomial regression analysis was started as follows with four main steps.

## Multinomial regression analysis

### Step 1: data preparation

First, the sample data set was imported into R, and the ordinal categorical variables (“Grade” and “Stage”) in the data were rewritten as ordered factors using the factor function. The level of the outcome to be used as the baseline was selected and specified using the relevel function. Here, “Alive” was defined as a baseline to be compared with “Dead (attributable to this cancer dx)” and “Dead of other cause” in the regression. The codes in R are shown below. The data we used in the analyses can be found in <http://cdn.amegroups.cn/static/application/949e8411be91d730d1670c07b0d01072/10.21037/atm-2020-57-1.pdf>.

```
## Import the data
seer_multi <- read.csv("D://Table S1.csv")
## Display part of the data
head(seer_multi)
tail(seer_multi)
## Rewrite the ordinal categorical variables
```

```
seer_multi$Grade <- factor(seer_multi$Grade, ordered = T,
  levels = c("GradeI", "GradeII",
  "GradeIII", "GradeIV"))
seer_multi$Stage <- factor(seer_multi$Stage, ordered = T,
  levels = c("I", "II", "III", "IV"))
## Specify the baseline level of the outcomes
seer_multi$Causespecificdeathclassification <- relevel(
  seer_multi$Causespecificdeathclassification, ref = "Alive")
```

### Step 2: model running

Second, the model was run using the multinom function in the nnet package in R as follows. In the outputs of the summary of the model (*Figure 1A*), a block of coefficients displayed as logged odds was shown followed by their standard errors.

Each of these blocks had one row of values corresponding to a model equation. In the first row, for instance, coefficients (logged odds) of each independent variable comparing “Dead (attributable to this cancer dx)” with “Alive” were shown. The coefficients were pulled out from the model and converted into interpretable odds ratios additionally by exponential analyses using the exp() command.

```
## Install and library the R package nnet
install.packages("nnet")
library(nnet)
## Run the multinomial model with the multinom function and
  summarize it
multimodel <- multinom(Causespecificdeathclassification ~
  AgeGroup + Sex + Race + PrimarySite +
  Grade + Stage + Chemotherapy,
  data = seer_multi)
summary(multimodel)
## Turn coefficients into interpretable odds ratios
oddsratio <- exp(coef(multimodel))
```

### Step 3: tests of the coefficients

After constructing the regression model, the next thing was to calculate P values of the regression coefficients using Wald tests (z tests used in this study). R codes and part of the outputs are shown below. The coefficients were considered to be of significance with a two-tailed value

```

A > summary(multimodel)
Call:
multinom(formula = Causespecificdeathclassification ~ AgeGroup +
  Sex + Race + PrimarySite + Grade + Stage + Chemotherapy,
  data = seer_multi)

Coefficients:
(Intercept) AgeGroup30-39 AgeGroup40-49 AgeGroup50-59 AgeGroup60-69 AgeGroup70-79
Dead (attributable to this cancer dx) -35.36552 34.20090 34.558119 34.650676 34.677491 34.875680
Dead of other cause -10.82465 -41.18071 8.139468 8.482122 8.959681 9.250254
AgeGroup80-89 AgeGroup90-99 SexMale RaceOther RaceWhite PrimarySiteMain
Dead (attributable to this cancer dx) 35.464585 36.63138 0.4905949 -0.4026380 -0.1380313 0.9148856
Dead of other cause 9.501117 10.39788 0.5636251 -0.5180268 -0.1538090 0.8668705
PrimarySiteMiddle PrimarySiteOverlapped PrimarySiteUpper Grade.L Grade.Q
Dead (attributable to this cancer dx) -0.4413495 0.2594643 -0.1571369 0.8478837 -0.05702993
Dead of other cause -0.7798977 -11.4127186 -0.2581571 0.8808246 0.08928033
Grade.C Stage.L Stage.Q Stage.C ChemotherapyYes
Dead (attributable to this cancer dx) -0.290519095 2.707031 0.1506986 -0.1339086 -1.048555
Dead of other cause 0.004185766 1.269906 -0.1179354 0.1475786 -1.612451

Std. Errors:
(Intercept) AgeGroup30-39 AgeGroup40-49 AgeGroup50-59 AgeGroup60-69 AgeGroup70-79
Dead (attributable to this cancer dx) 0.2154735 6.288912e-01 0.2411192 0.1548453 0.1425103 0.1466400
Dead of other cause 0.3609044 1.872785e-16 0.6393291 0.2943932 0.2270343 0.2277458
AgeGroup80-89 AgeGroup90-99 SexMale RaceOther RaceWhite PrimarySiteMain
Dead (attributable to this cancer dx) 0.1680675 0.5099957 0.09611013 0.2020646 0.1487253 0.3367331
Dead of other cause 0.2610726 0.7259945 0.18983199 0.4313394 0.2975992 0.6029937
PrimarySiteMiddle PrimarySiteOverlapped PrimarySiteUpper Grade.L Grade.Q Grade.C
Dead (attributable to this cancer dx) 0.2382936 4.407891e-01 0.1072204 0.3104837 0.2356134 0.1239088
Dead of other cause 0.5437503 8.360122e-07 0.2032950 0.5612626 0.4265672 0.2293663
Stage.L Stage.Q Stage.C ChemotherapyYes
Dead (attributable to this cancer dx) 0.1206860 0.1437696 0.1701652 0.1122830
Dead of other cause 0.1948825 0.2356804 0.2789157 0.2414103

Residual Deviance: 3650.35
AIC: 3738.35

B > anova(multimodelfit1, multimodelfit2, multimodelfit3, multimodelfit4, multimodelfit5, multimodelfit6,
+ multimodelfit7, multimodelfit8)
Likelihood ratio tests of Multinomial Models

Response: Causespecificdeathclassification

Model Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 1 5998 4868.684
2 AgeGroup 5984 4817.917 1 vs 2 14 50.766558 4.540814e-06
3 AgeGroup + Sex 5982 4770.538 2 vs 3 2 47.378950 5.149814e-11
4 AgeGroup + Sex + Race 5978 4763.119 3 vs 4 4 7.419143 1.153281e-01
5 AgeGroup + Sex + Race + PrimarySite 5970 4711.442 4 vs 5 8 51.676712 1.943170e-08
6 AgeGroup + Sex + Race + PrimarySite + Grade 5964 4482.810 5 vs 6 6 228.632575 0.000000e+00
7 AgeGroup + Sex + Race + PrimarySite + Grade + Stage 5958 3765.321 6 vs 7 6 717.488894 0.000000e+00
8 AgeGroup + Sex + Race + PrimarySite + Grade + Stage + Chemotherapy 5956 3650.350 7 vs 8 2 114.971331 0.000000e+00

```

**Figure 1** Outputs of the multinomial logistic regression model. (A) Summary of the model. (B) Results of ANOVA. Grade. L: Grade II; Grade. Q: Grade III; Grade. C: Grade IV; Stage. L: Stage II; Stage. Q: Stage III; Stage. C: Stage IV.

of  $P < 0.05$ , and no significance was attributed otherwise. Focusing on the block of P values below, variables with significant P values of their coefficients were determined as significant prognostic factors that contributed significantly differently to cancer-specific death or death of other causes rather than to survival in patients with lung adenocarcinoma.

```
## Wald tests/z tests of the coefficients
```

```

zvalues <- summary(multimodel)$coefficients /
summary(multimodel)$standard.errors
pvalues <- pnorm(abs(zvalues), lower.tail = FALSE) * 2
pvalues

```

```

## (Intercept) AgeGroup30-39
## Dead (attributable to this cancer dx) 0.000000e+00 0
## Dead of other cause 1.206239e-197 0
...

```

```
## PrimarySiteUpper Grade.I
##Dead (attributable to this cancer dx) 0.1427709 0.006317171
##Dead of other cause 0.2041328 0.116563433
...
## ChemotherapyYes
##Dead (attributable to this cancer dx) 9.770398e-21
##Dead of other cause 2.400872e-11
```

Combining the results of the logged odds (coefficients) in *Figure 1A* with those of the P values in this part, the coefficients could be interpreted as changes in odds for a certain category comparing with another category in each independent variable. For example, the logged odds of cancer-specific death due to lung adenocarcinoma versus survival significantly increased by approximately 0.848 if moving from differentiation “Grade I” to “Grade II” of tumors with the logged odds being 8.478837e-1 and the P value being 6.317171e-3.

#### Step 4: tests and validations of the model

Last but not least, the analysis of variance (ANOVA) was used to test the significance of this regression model using the ANOVA function. Using this function, the goodness of fit of the model was assessed indirectly by comparing regression models with the regression models before adding variables. Here, significance was considered with a relaxed value of  $P < 0.10$  or  $P < 0.15$  when comparing multiple regression models. The results are shown in *Figure 1B*. The P value in the last row (model 1 vs. model 8) indicated that the proposed multinomial regression model was of significance. Besides, the second to the eighth models were all significant improvements over previous models because their P values were generally less than 0.15.

```
## Using the ANOVA to test the model
multimodelfit1 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ 1)
multimodelfit2 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup)
multimodelfit3 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex)
multimodelfit4 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex + Race)
```

```
multimodelfit5 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex + Race + PrimarySite)
multimodelfit6 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex + Race + PrimarySite + Grade)
multimodelfit7 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex + Race + PrimarySite + Grade +
  Stage)
multimodelfit8 <- multinom (data = seer_multi,
  Causespecificdeathclassification ~ AgeGroup +
  Sex + Race + PrimarySite + Grade +
  Stage + Chemotherapy)
anova(multimodelfit1, multimodelfit2, multimodelfit3,
  multimodelfit4, multimodelfit5, multimodelfit6,
  multimodelfit7, multimodelfit8)
```

Additionally, the regression model was further validated by predicting the survival outcomes of each patient in the data set together with the probabilities of each outcome using the predict function in R as follows. The prediction accuracy of the model was obtained by calculating the mean of predicted probabilities of the original survival outcomes of each patient. The results demonstrated that 74.5% of the clinical survival outcomes of the patients in the present data set were verified to be correct using the regression model, indicating that this multinomial regression model had strong reliability.

```
## Predict the category of outcomes for each patient
pred.multinom <- predict(multimodel, seer_multi)
head(pred.multinom)

##[1] Dead (attributable to this cancer dx)
##[2] Dead (attributable to this cancer dx)
##[3] Alive
...

## Probabilities with each outcome for each patient
pprob <- predict(multimodel, seer_multi, type = "p")
head(pprob)

##Alive Dead (attributable to this cancer dx) Dead of other
cause
```

```
##1 0.1166890 0.81108660 0.07222440
##2 0.1237749 0.80505403 0.07117102
##3 0.9122983 0.05056706 0.03713464
...

## Calculate the prediction accuracy of the model
pred_accuracy <- mean(pred.multinom == seer_multi$
  Causespecificdeathclassification)
pred_accuracy

##[1] 0.7453333
```

### Ordinal logistic regression model

Next, an ordinal logistic regression was displayed similarly using the R software with the sample data. Here, five steps in total should be taken in constructing an ordinal logistic regression model as follows.

#### Step 1: data preparation

This step was basically the same as the processes in the first step of multinomial regression analysis, including data import and variable redefinition. Relevant R codes can refer to the previous section of multinomial regression.

#### Step 2: model running

The ordinal regression model was run using the `polr` function in the MASS package of R. The codes are shown as follows. The summary of the model (Figure 2A) also contained a block of coefficients displayed as logged odds, followed by their standard errors together with t values. These coefficients could also be converted into interpretable odds ratios using the `exp()` command.

Each row of values corresponded to a model equation with a category of a certain independent variable. The coefficients in each row were indicators of the logged times the patient survival will increase by one degree (“0 = no response” to “1 = partial response” or “1 = partial response” to “2 = complete response”) when one of their clinicopathological characteristics changes from the reference category to a certain category displayed in the beginning of the row.

```
## Install and library the R package MASS
```

```
install.packages("MASS")
library(MASS)
## Run the ordinal model with polr function and summary it
ordinalmodel <- polr(data = seer_ordinal, SurvivalStatus ~
  AgeGroup + Sex + Race + PrimarySite +
  Grade + Stage + Chemotherapy)
summary(ordinalmodel)
## Turn coefficients into interpretable odds ratios
oddsratio2 <- exp(coef(ordinalmodel))
```

#### Step 3: test of parallel lines

Parallel regression assumption or the proportional odds assumption is a necessity for the application of the ordinal logistic regression model for an ordered categorical variable; otherwise, the multinomial model described earlier has to be used. This assumption can be tested using a Brant test in the R software, which is available in the Brant package with the `brant` function. A P value higher than 0.05 may lead to the failure of rejecting the null hypothesis that the parallel regression assumption holds or that the coefficients do not differ across different cut points in the outcome variable. According to the codes and outputs, there was a P value of 0.67 in the “Omnibus” check, indicating the lack of evidence showing that the previous assumption has not been met. Therefore, the ordinal regression method was appropriate for the sample data set.

```
## Test of parallel lines
install.packages("brant")
library(brant)
brant(ordinalmodel)

##-----
##Test for → → → X2 → df → probability
##-----
##Omnibus → → → 184.08 → 21 → 0.67
##AgeGroup30-39 → → → 0 → 1 → 0.99
## ...
##-----
##H0: Parallel Regression Assumption holds
```

#### Step 4: test of the coefficients

Different from the multinomial regression model, the

A

```
> summary(ordinalmodel)

Re-fitting to get Hessian

Call:
polr(formula = SurvivalStatus ~ AgeGroup + Sex + Race + PrimarySite +
      Grade + Stage + Chemotherapy, data = seer_ordinal)

Coefficients:
                Value Std. Error  t value
AgeGroup30-39  -0.23719   1.67874  -0.14129
AgeGroup40-49   0.15718   1.60014   0.09823
AgeGroup50-59   0.14791   1.59114   0.09296
AgeGroup60-69   0.17564   1.59027   0.11045
AgeGroup70-79   0.13736   1.59069   0.08635
AgeGroup80-89  -0.04981   1.59239  -0.03128
AgeGroup90-99  -0.96354   1.64408  -0.58606
SexMale        -0.21489   0.07196  -2.98603
RaceOther      -0.01993   0.15643  -0.12742
RaceWhite      0.08750   0.11434   0.76523
PrimarySiteMain -0.51322   0.27678  -1.85426
PrimarySiteMiddle 0.12928   0.17337   0.74569
PrimarySiteOverlapped -0.27303  0.36712  -0.74371
PrimarySiteUpper 0.15623   0.08014   1.94943
Grade.L        -0.56152   0.24686  -2.27467
Grade.Q         0.23328   0.18649   1.25087
Grade.C         0.16276   0.09675   1.68236
Stage.L        -1.77216   0.08322 -21.29373
Stage.Q        -0.25581   0.09716  -2.63305
Stage.C         0.05645   0.11456   0.49276
ChemotherapyYes 1.18624   0.09116  13.01301

Intercepts:
      Value  Std. Error t value
0|1  -0.3015   1.5990   -0.1885
1|2   1.6309   1.5992   1.0199

Residual Deviance: 5718.301
AIC: 5764.301
```

```
B > anova(ordinalmodelfit1, ordinalmodelfit2, ordinalmodelfit3, ordinalmodelfit4,
+        ordinalmodelfit5, ordinalmodelfit6, ordinalmodelfit7, ordinalmodelfit8)
Likelihood ratio tests of ordinal regression models
```

Response: SurvivalStatus		Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
1		1	2998	6558.974				
2		AgeGroup	2991	6534.652	1 vs 2	7	24.322453	9.997700e-04
3		AgeGroup + Sex	2990	6512.360	2 vs 3	1	22.292199	2.341559e-06
4		AgeGroup + Sex + Race	2988	6508.533	3 vs 4	2	3.827132	1.475533e-01
5		AgeGroup + Sex + Race + PrimarySite	2984	6476.864	4 vs 5	4	31.669150	2.235288e-06
6		AgeGroup + Sex + Race + PrimarySite + Grade	2981	6313.402	5 vs 6	3	163.461900	0.000000e+00
7		AgeGroup + Sex + Race + PrimarySite + Grade + Stage	2978	5898.679	6 vs 7	3	414.722401	0.000000e+00
8		AgeGroup + Sex + Race + PrimarySite + Grade + Stage + Chemotherapy	2977	5718.301	7 vs 8	1	180.378453	0.000000e+00

**Figure 2** Outputs of the ordinal logistic regression model. (A) Summary of the model. (B) Results of ANOVA. Grade. L: Grade II; Grade. Q: Grade III; Grade. C: Grade IV; Stage. L: Stage II; Stage. Q: Stage III; Stage. C: Stage IV.

ordinal regression model has provided the t values of each of the categories of variables that can be directly used to calculate the P values of the coefficients by Wald tests using the following codes. Combining the results of logged odds (coefficients) and their P values, the coefficients can be interpreted as changes in odds for a certain category of

variables, but the interpretation is slightly different from the previous multinomial logistic regression. For example, if moving from differentiation “Grade I” to “Grade II” of tumors, the logged odds of the probability of patients being in the next degree of survival time (from “0 = no response” to “1 = partial response” or from “1 = partial response” to

“2 = complete response”) will be decreased by  $-0.562$  with significance.

```
## Wald tests/z tests of the coefficients
tvalue <- btotal2$`t value`
pvalue <- pnorm(abs(tvalue), lower.tail = FALSE) * 2
pvalue

## [1] 8.877339e-01 ##AgeGroup30-39
...
##[15] 3.041001e-03 ##GradeGradeII
...
##[21] 1.032186e-38 ##ChemotherapyYes
```

### Step 5: test and validations of the model

Same as the fourth step in multinomial regression, the significance of the ordinal regression model and the goodness of fit of the proposed model needs to be assessed using the ANOVA function. R codes can refer to the multinomial regression analysis section. Results (Figure 2B) indicated that the proposed ordinal regression model was significant, and the second to eighth models were validated to be significant improvements over each of their previous ones.

As in a multinomial regression model, further validations can be performed using the predict function. R codes can be found in the same step of the previous section. Here, our proposed multinomial regression model had relatively strong reliability since the accuracy was shown to be 52.2%.

### Conclusions

In summary, two logistic regression methods were introduced in this study using the R software for multi-categorical variables in four basic steps of statistical regression: data preparation, model establishment, tests of the coefficients and the model, and model validations. The authors are grateful to Dr. Zhou and his colleagues for their contributions, which served as a valuable reference.

### Acknowledgments

The authors thank the International Science Editing Co. (<http://www.internationalscienceediting.com>) for editing this manuscript.

*Funding:* None.

### Footnote

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-2020-57>). CZ serves as an unpaid Section Editor of *Annals of Translational Medicine* from Oct 2019 to Sep 2020. The other authors have no conflicts of interest to declare.

*Ethical statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Ethics approval has been exempted by the Ethics Committee of Zhongshan Hospital of Fudan University (Shanghai, China), as the SEER is a publicly available database, and data extracted from SEER are identified as nonhuman study.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

### References

1. LaValley MP. Logistic regression. *Circulation* 2008;117:2395-9.
2. Kwak C, Clayton-Matthews A. Multinomial logistic regression. *Nurs Res* 2002;51:404-10.
3. Zhou ZR, Wang WW, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. *Ann Transl Med* 2019;7:796.
4. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925-31.
5. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical prediction models involving radiological parameters: study design and statistical methods. *Korean J Radiol* 2016;17:339-50.
6. Woodward M, Tunstall-Pedoe H, Peters SA. Graphics and statistics for cardiology: clinical prediction rules. *Heart* 2017;103:538-45.
7. Bi G, Li R, Liang J, et al. A nomogram with

- enhanced function facilitated by `nomogramEx` and `nomogramFormula`. *Ann Transl Med* 2020;8:78.
8. Tolles J, Meurer WJ. Logistic regression: Relating patient characteristics to outcomes. *JAMA* 2016;316:533-4.

9. García-Pérez MA. Statistical criteria for parallel tests: a comparison of accuracy and power. *Behav Res Methods* 2013;45:999-1010.

**Cite this article as:** Liang J, Bi G, Zhan C. Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R. *Ann Transl Med* 2020;8(16):982. doi: 10.21037/atm-2020-57