



ORIGINAL ARTICLE

A nomogram was developed to enhance the use of multinomial logistic regression modeling in diagnostic research

Loes C.M. Bertens*, Karel G.M. Moons, Frans H. Rutten, Yvonne van Mourik, Arno W. Hoes, Johannes B. Reitsma

Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, PO Box 85060, Stratenum 6.131, Utrecht 3508 AB, The Netherlands

Accepted 28 October 2015; Published online xxxx

Abstract

Objectives: We developed a nomogram to facilitate the interpretation and presentation of results from multinomial logistic regression models.

Study Design and Setting: We analyzed data from 376 frail elderly with complaints of dyspnea. Potential underlying disease categories were heart failure (HF), chronic obstructive pulmonary disease (COPD), the combination of both (HF and COPD), and any other outcome (other). A nomogram for multinomial model was developed to depict the relative importance of each predictor and to calculate the probability for each disease category for a given patient. Additionally, model performance of the multinomial regression model was assessed.

Results: Prevalence of HF and COPD was 14% ($n = 54$), HF 24% ($n = 90$), COPD 20% ($n = 75$), and Other 42% ($n = 157$). The relative importance of the individual predictors varied across these disease categories or was even reversed. The pairwise C statistics ranged from 0.75 (between HF and Other) to 0.96 (between HF and COPD and Other). The nomogram can be used to rank the disease categories from most to least likely within each patient or to calculate the predicted probabilities.

Conclusions: Our new nomogram is a useful tool to present and understand the results of a multinomial regression model and could enhance the applicability of such models in daily practice. © 2015 Elsevier Inc. All rights reserved.

Keywords: Multinomial; Regression analyses; Nomogram; Heart failure; Chronic obstructive Pulmonary disease; Diagnostic research

1. Introduction

The starting point in clinical practice is a patient with certain symptoms and/or signs, and nearly always, different disorders can be responsible. This is known as the differential diagnosis. Additionally, several disorders can be present simultaneously in a single patient. In diagnostic research, the presence or absence of a single disease is usually modeled using dichotomous logistic regression. In such models, there is one disease of interest and patients with no or alternative diseases are combined into the disease

absent category. Consequently, the multinomial aspect of clinical diagnosis is ignored [1,2]. To mimic the diagnostic process in real-life clinical practice more closely, multinomial logistic regression modeling could be applied, considering multiple diseases and their combination as potential outcomes simultaneously [3–7].

Multimorbidity is common in the elderly, and diseases often result in overlapping symptoms and signs. Patients presenting with dyspnea form a typical example. More than 30 diseases can be the underlying cause; in the elderly, chronic obstructive pulmonary disease (COPD) and heart failure (HF) alone or in combination are the most likely explanation of dyspnea [8–10]. When multiple disorders may be present in a single patient, multinomial logistic regression modeling seems an attractive method in diagnostic research, as it models estimates for all outcomes of interest simultaneously.

In the previously mentioned situation, the use of a multinomial regression model seems more natural than the use of multiple dichotomous regression models. However, multinomial regression modeling is not frequently applied in clinical research, despite the statistical advantages of using such

Funding: The study was conducted as part of the Dutch National Care for the Elderly Program (ZonMw-NPO) and funded by a research grant from the “Netherlands Organization for Health Research and Development” (ZonMw grant 311040302) and a grant from the Netherlands Organization for Scientific Research (project 918.10.615).

Conflict of interest: None.

Trial registration: ClinicalTrials.gov NCT01148719.

* Corresponding author. Tel.: +31-88-756-9415; fax: +31-88-756-9028.

E-mail address: L.C.M.Bertens-2@umcutrecht.nl or <http://www.juliuscenter.nl/L.C.M.Bertens>.

<http://dx.doi.org/10.1016/j.jclinepi.2015.10.016>

0895-4356/© 2015 Elsevier Inc. All rights reserved.

What is new?**Key findings**

- The multinomial nomogram helps visualizing the relative importance of predictors for the different disease categories and assists in the calculation of absolute probabilities for the various outcome categories for a specific patient.

What this adds to what was known?

- We developed a nomogram for multinomial models to facilitate the interpretation and presentation of the results from multinomial regression models.

What is the implication and what should change now?

- The multinomial model should be used more in diagnostic research, especially in situations where multiple diseases are considered including the likelihood that a combination of those diseases is present.

models [4,5]. One possible reason is that the presentation and calculation of predicted probabilities from a multinomial logistic regression model is more complex than from a binary regression model. Nomograms are increasingly being used in binary logistic regression models and Cox survival models to facilitate the interpretation of their results [11]. Given the more complex interpretation of regression coefficients from multinomial logistic regression models, the usefulness of nomograms is likely to be higher.

Therefore, we aim to improve the use of multinomial regression models for clinical practice by describing a nomogram to present the results of a multinomial logistic regression model. We will use data from a diagnostic study in frail elderly patients with dyspnea and/or reduced exercise tolerance who were examined for the presence of HF, COPD, or both [12,13].

2. Methods

2.1. Case study

2.1.1. Participants

The study population was derived from a cluster randomized trial in which community-dwelling frail elderly with complaints of dyspnea and/or reduced exercise tolerance were evaluated (triage of reduced exercise tolerance in frail elderly) [12,13]. In this study, frailty was defined as three or more comorbidities or the chronic use of five or more drugs. For the present study, we used data from those randomized to the screening arm of the trial. All these participants underwent a diagnostic strategy, including history taking, physical examination, electrocardiography,

spirometry, blood tests, and echocardiography. The study complied with the Declaration of Helsinki, and the Medical Ethical Committee of the University Medical Centre Utrecht approved the study (trial registration: ClinicalTrials.gov NCT01148719). All participants gave their written informed consent.

2.1.2. Outcome

The outcome consisted of four categories: COPD alone, HF alone, the combination of HF and COPD (HF and COPD), and another or no disease (Other). These final diagnoses were established by a panel of experts during a consensus meeting. The panel always consisted of a general practitioner (F.H.R.), a pulmonologist (alternating: J.W.J.L. or H.J.H.), and a cardiologist (alternating: M.J.M.C., M.A.N.S., or C.G.K.M.F.). Signs, symptoms, and test results from the diagnostic strategy, including spirometry and echocardiography, from each patient were discussed before reaching a consensus decision on the presence or absence of a particular diagnosis.

2.1.3. Potential diagnostic predictors

Based on the literature, nine variables were selected as potential diagnostic predictors for the presence of HF and/or COPD: gender, body mass index (BMI), signs of fluid overload (a composite of peripheral edema, pulmonary crepitations, nocturnal dyspnea, orthopnea, and elevated jugular venous pressure), displaced apex beat, NT-proBNP levels, pack years of smoking, breathing sounds (a composite of wheezing on history and physical examination and rhonchi), cough, and forced expiratory volume in 1 second as percentage of the population-specific predicted values (FEV1) [9,14–19]. BMI, number of pack years of smoking, NT-proBNP, and FEV1 were treated as continuous variables.

2.2. Missing data

Overall, very few values were missing in our data set. In three patients, the spirometry data were unreliable due to poor technical performance, and presence or absence of COPD could not be determined in these patients. In 10 patients, echocardiography was missing, and presence or absence of HF could not be determined. For the present study, these 13 patients were excluded. Of the determinants, 17 values of NT-pro-BNP were missing, and six missing values were of postbronchodilator FEV1 measurements. These values were imputed using multiple imputation (10 rounds of imputation), and Rubin's rules were applied to come to the overall estimates in the regression models [20,21].

2.3. Multinomial logistic regression model

A multinomial logistic regression model was fitted in which the probability of each of the four disease categories (HF and COPD, HF, COPD, and Other) was estimated in a single model using maximum likelihood techniques [4,7].

For model development, backward selection procedures applying a P -value <0.10 for the log-likelihood ratio test were used to select the predictors in the model. In a multinomial logistic model, a single algorithm is used to estimate the regression coefficients of the individual predictors for each disease category simultaneously, thereby enabling different estimates of a single predictor across the disease categories HF and COPD, HF, and COPD. The disease category “Other” was used as the reference category in our model. The resulting combinations of intercept and coefficients were used to calculate the three linear predictors (LPs). The predicted probabilities for each category can be calculated with these LPs. For instance, the predicted probability for the presence of both HF and COPD is calculated as follows: $\text{probability}_{\text{HF and COPD}} = \exp(\text{lp}_{\text{HF and COPD}}) / (1 + \exp(\text{lp}_{\text{HF and COPD}}) + \exp(\text{lp}_{\text{HF}}) + \exp(\text{lp}_{\text{COPD}}))$.

The predicted probabilities for (only) HF and (only) COPD were calculated in the same way. For the reference category “Other,” for which no LP is available, the predicted probability was calculated by subtracting the three calculated predicted probabilities from 1.

For each individual, the predicted disease category was identified as the category with the highest predicted probability of all four categories. Classification tables were assembled by cross-classifying the observed diagnostic categories (i.e., panel diagnoses) by the predicted disease categories from the multinomial model. The ability of the multinomial logistic regression model to discriminate between the four diagnostic categories was quantified by calculating pairwise C statistics between each disease category, for example between HF and COPD and Other [22,23].

2.4. Multinomial nomogram

The results of the multinomial logistic regression model were presented using a nomogram for disease categories COPD, HF, and HF and COPD. Each horizontal line in the nomogram shows the relative effect of the predictors on the different disease categories compared with the reference category (Other). Longer lines represent higher scores and are indicative of larger effects of the predictor coefficient on that particular disease category. For each patient, disease-specific scores can be “read” of the nomogram corresponding to different patient characteristics. The sum of these predictor scores provides insight into the relative likelihood that a patient with those characteristics has COPD, HF, HF and COPD, or none of these. “Relative” means that the sum score will update the prior probabilities of each disease category, with higher scores resulting in larger changes in probabilities. Therefore, to determine the absolute probabilities for each disease category, the prior probability needs to be taken into account.

To obtain easy-to-sum scores, all scores corresponding to individual predictors were rescaled to values between 0 and 100 in the following way. First, the ranges of the

continuous variables (BMI, pack years, NT-proBNP, and FEV1) were adjusted to exclude outliers before the nomogram was fitted. The new ranges were set to 20 to 40 for BMI, 0 to 100 for pack years, 10 to 4,000 for NT-proBNP, and 40 to 120 for FEV1. Original values outside these new ranges were truncated to the min or max of these ranges. Next, the highest possible score was calculated for each predictor coefficient for each disease category (predictor–coefficient pair, e.g., cough_{HF}). This score was calculated by multiplying the coefficients with the maximum value of the corresponding predictor. The predictor with the highest possible score was identified and a maximum score of 100 points was assigned to the maximum value of this predictor, and all other predictor values were rescaled using the same scaling factor. Finally, for predictors with negative model coefficients, the values on the nomogram were placed in the opposite direction. For example, the predictor BMI, in case of a negative model coefficient, the range of values on the nomogram goes from 40 to 20, instead of from 20 to 40. A more detailed description of the nomogram development can be found in Appendix A at www.jclinepi.com. The scores for the different disease categories can be used to determine the most likely disease category for that patient by transforming the scores into LPs and then calculate the predicted probabilities. Details of these transformations and a worked out example can be found in Appendix B at www.jclinepi.com. The R code used to develop the multinomial nomogram is available on request.

The analyses were performed with R version 3.0.3 (2014, The R Foundation for Statistical Computing), packages *nnet* and *rms*.

3. Results

The baseline characteristics of the 376 study patients in each of the four disease categories (HF and COPD, HF, COPD, and other) are shown in Table 1. The combination of HF and COPD was diagnosed in 14% ($n = 54$), HF alone in 24% ($n = 90$), COPD alone in 20% ($n = 75$), and the remaining category “Other” in 42% ($n = 157$) of the participants.

The final multinomial logistic regression model, in which disease categories COPD, HF, and HF and COPD were modeled against disease category Other as reference, included eight variables (Table 2). Several predictor variables had different associations with the disease categories. For example, the odds ratio (OR) for “displaced apex beat” differed markedly in size between disease categories, with an OR of 1.4 for COPD, 4.0 for HF, and 10.6 for HF and COPD.

3.1. Discrimination and classification

The multinomial logistic regression model showed good discriminative ability with pairwise C statistics ranging from 0.75 (between HF and Other) to 0.96 (between HF

Table 1. Demographic and clinical characteristics of 376 frail elderly divided into four patient groups based on the final panel diagnoses being present: (only) COPD, (only) HF, HF and COPD, and Other

Determinants	COPD (n = 75)	HF (n = 90)	HF and COPD (n = 54)	Other (n = 157)
Male (%)	38 (50.7)	45 (50.0)	27 (50.0)	62 (39.5)
Mean BMI (±SD)	27.1 (3.6)	29.1 (4.6)	28.7 (5.0)	28.1 (4.3)
Median pack years, (IQR)	24.0 (0.0–42.0)	3.4 (0.0–22.5)	19.0 (0.0–35.6)	5.0 (0.0–20.0)
Signs of fluid overload (%)	38 (50.7)	64 (71.1)	46 (85.2)	76 (48.4)
Displaced apex beat (%)	9 (12.0)	19 (21.1)	18 (33.3)	12 (7.6)
Abnormal breathing sounds (%)	35 (46.7)	20 (22.2)	33 (61.1)	28 (17.8)
Cough (%)	33 (44.0)	19 (21.1)	8 (14.8)	28 (17.8)
Median NT-proBNP (pg/mL), (IQR) ^a	137.5 (94.0–217.5)	282.0 (125.8–713.3)	347.5 (185.3–855.3)	99.0 (58.0–187.0)
FEV1, as % predicted (±SD) ^b	74.2 (16.9)	95.3 (15.3)	72.5 (14.0)	99.9 (17.2)

Abbreviations: HF, heart failure; COPD, chronic obstructive pulmonary disease; HF and COPD, the combination of HF and COPD; BMI, body mass index; IQR, interquartile range; SD, standard deviation; signs of fluid overload, edema, crepitations, nocturnal dyspnea, orthopnea, and elevated jugular venous pressure; abnormal breathing sounds, wheezing on history and physical examination and rhonchi; FEV1, postbronchodilator forced expiratory volume in 1 second as % predicted.

^a n = 359 (COPD: 72, HF: 86, HF–COPD: 50, and Other: 151).

^b n = 371 (COPD: 73, HF: 89, HF–COPD: 54, and Other: 155).

and COPD and Other) and are displayed in Fig. 1. The multinomial model classified in total 63.8% of patients correctly.

3.2. Multinomial nomogram

The multinomial nomogram presented in Fig. 2 can be used to present the prediction model (Table 2) more comprehensively. The different lines in the nomogram can be interpreted as the different effect sizes of the model coefficients, with longer lines (and thus higher scores) representing larger effects. How to use the nomogram to calculate the scores for COPD, HF, and HF and COPD is explained below with an example.

Consider a patient with a BMI of 35.4 kg/m², a history of 48 pack years of cigarette smoking, NT-proBNP serum

levels of 981 pg/mL, FEV1 of 89% of predicted, and a displaced apex beat.

The patient characteristics can be converted into scores for HF and COPD as follows: First, determine the position of 35.4 on the HF and COPD line for BMI. Then, draw a vertical line to the upper line indicated by “points.” The number of points for BMI is where these lines cross. In this case, the patient receives 11 points for his BMI when considering a diagnosis of HF and COPD. Now, the same is done for the other predictors for this disease, resulting in: 15 points for pack years, 75 points for NT-proBNP, 38 points for FEV1, no points for signs of fluid overload and abnormal breathing sounds, 31 points for displaced apex beat, and 10 points for the absence of cough. Summing these points resulted in a total score of 180 points for HF and COPD in this patient. These steps can be repeated to

Table 2. Multivariable estimates from the final multinomial logistic regression model

Diagnostic determinants	COPD vs. Other		HF vs. Other		HF and COPD vs. Other	
	Coefficients	OR (95% CI)	Coefficients	OR (95% CI)	Coefficients	OR (95% CI)
Intercept	6.84		−9.14		−3.89	
BMI	−0.05	0.95 (0.87–1.04)	0.07	1.07 (1.00–1.15)	0.06	1.06 (0.96–1.17)
Pack years ^a	0.69	1.99 (1.21–3.29)	0.08	1.08 (0.70–1.68)	0.67	1.95 (1.06–3.59)
NT-proBNP ^a	0.56	1.75 (0.74–4.15)	2.59	13.33 (6.09–29.20)	2.87	17.64 (6.26–49.66)
FEV1, % predicted	−0.10	0.90 (0.88–0.93)	−0.00	1.00 (0.98–1.02)	−0.09	0.91 (0.89–0.94)
Signs of fluid overload	−0.64	0.53 (0.25–1.11)	0.73	2.08 (1.08–3.99)	1.09	2.97 (1.06–8.34)
Displaced apex beat	0.34	1.40 (0.43–4.56)	1.39	4.01 (1.62–9.93)	2.36	10.59 (3.40–33.02)
Abnormal breathing sounds	0.63	1.88 (0.86–4.11)	0.38	1.46 (0.68–3.14)	1.71	5.53 (2.13–14.36)
Cough	1.28	3.60 (1.61–8.03)	0.31	1.36 (0.63–2.93)	−0.74	0.48 (0.15–1.49)

Abbreviations: COPD, chronic obstructive pulmonary disease; HF, heart failure; HF and COPD, the combination of HF and COPD; OR, odds ratio; CI, confidence interval; BMI, body mass index; LP, linear predictor.

Significant ($P < 0.05$) associations are depicted in bold.

Diagnostic determinant gender showed $P > 0.10$ during backward selection procedures and was therefore excluded from the final model.

For example: consider a patient with a BMI of 35.4 kg/m², history of 48 pack years, NT-proBNP serum levels of 981 pg/mL, FEV1 of 89% predicted and a displaced apex beat. Disease probabilities are calculated using the following equation: probability COPD = exp (LP(COPD)/ (1 + exp (LP(COPD)) + exp (LP(HF)) + exp (LP(HF and COPD))). The LP of COPD is (6.835 − 0.048 × 35.4 + 0.689 × log(48) + 0.563 × log(981) − 0.095 × 89 + 0.337) = −0.14. The LP of HF is (−9.139 + 0.073 × 35.4 + 0.082 × log(48) + 2.589 × log(981) − 0.001 × 89 + 1.387) = 2.63. The LP of HF and COPD is (−3.891 + 0.055 × 35.4 + 0.671 × log(48) + 2.867 × log(981) − 0.089 × 89 + 2.364) = 2.21. The disease probability of COPD is 0.04; 0.56 for HF and 0.37 for HF and COPD and the probability of Other is 1 − 0.04 − 0.56 − 0.37 = 0.03. Resulting in HF as the predicted outcome category.

^a values are log transformed.

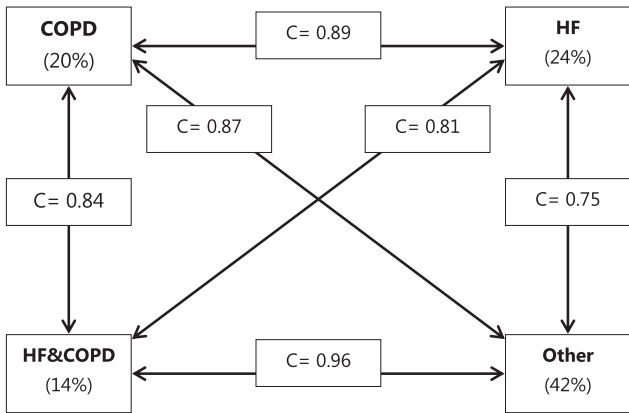


Fig. 1. Schematic presentation of the discriminative ability of the multinomial regression model. The figure shows the prevalence of each disease category (%) as well as the pairwise C statistic (C) between the different disease categories. HF, heart failure; COPD, chronic obstructive pulmonary disease; HF and COPD, the combination of HF and COPD; Other, all other diagnoses including no disease.

calculate the total scores for COPD (85 points) and HF (103 points).

These total scores can be used to calculate the corresponding predicted probabilities (as described in Appendix B at www.jclinepi.com); the predicted probability for HF and COPD is 0.38, for COPD 0.04, for HF 0.54, and for disease category Other 0.04.

4. Discussion

In this article, we introduced and explained a novel type nomogram for the presentation and understanding of a multinomial logistic regression model.

The ORs and intercepts from a multinomial logistic regression model are interpreted similarly as those from dichotomous logistic regression models. In a multinomial model, these ORs and intercepts are estimated directly in a single multinomial regression model, which enables an easy comparison of the estimates for each disease category, and unexpected differences are more easily detected than when separate dichotomous models are applied. This may be of specific interest when unexpected differences are observed between the diagnostic value for certain predictors for the disease category representing the patients with both diseases (HF and COPD) and the individual disease outcome COPD or HF. An example is the predictor variable “cough,” with an OR of 0.48 for HF and COPD and the ORs for COPD and HF being larger than 1.0 (OR for COPD of 3.60 and for HF 1.36).

Additionally, we developed a nomogram to facilitate the presentation and understanding of the multinomial logistic regression model. The different lines per predictor in the nomogram visualize the differences in relative importance of association (i.e., size of coefficients) for each disease category, thereby making the interpretation of the

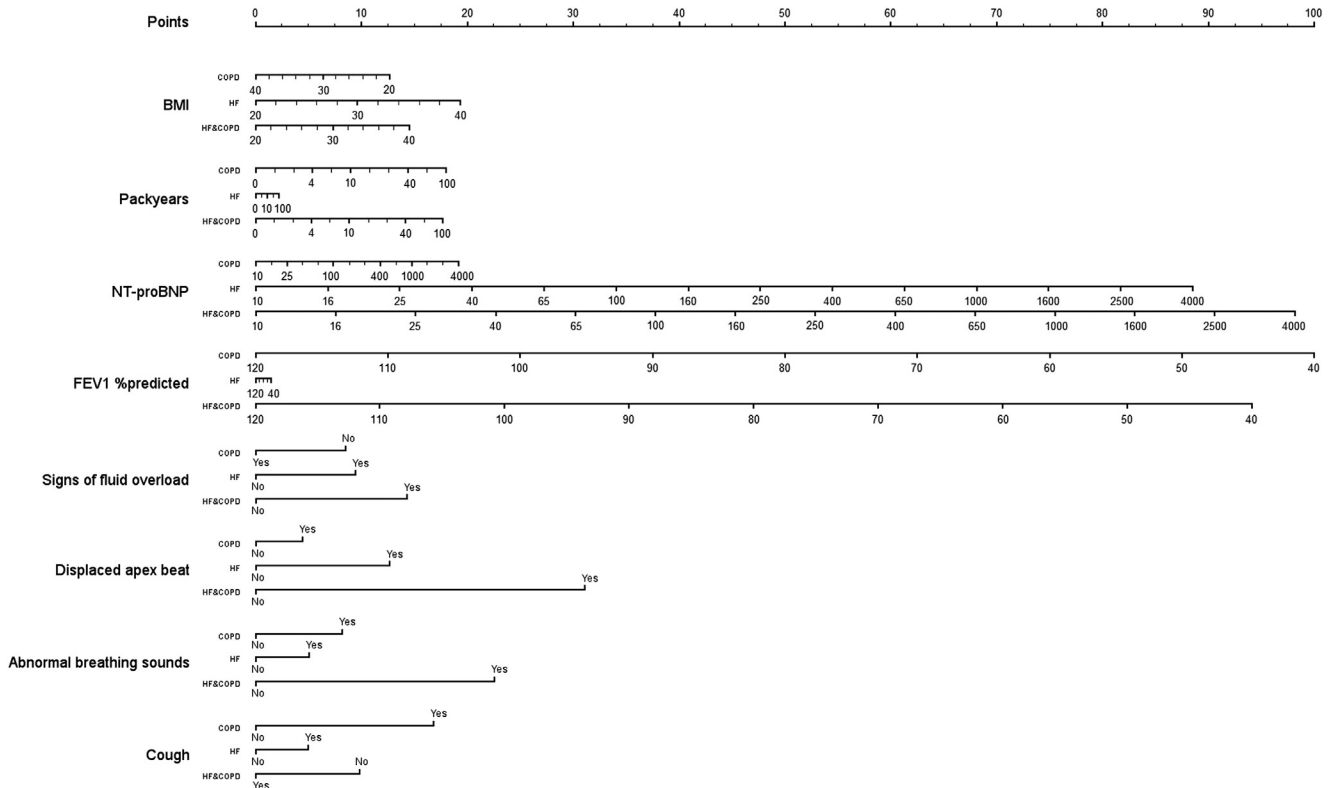


Fig. 2. Multinomial nomogram for the prediction of Heart failure, COPD and the combination of both diseases. BMI, body mass index; FEV1, forced expiratory volume in 1 second as % predicted.

differences in regression coefficients of the predictors per disease categories more easy in this nomogram than in the more traditional presentation of a (multinomial) regression model, in a table showing all ORs. This nomogram can also be used to calculate the predicted probabilities as the total scores are directly related to the model coefficients for each disease category. To calculate predicted probabilities, the total scores need to be combined with the different intercepts (see Appendix B at www.jclinepi.com). The predicted probabilities derived from the multinomial nomogram can be slightly different from the probabilities calculated directly from the model due to rounding and reading off the nomogram scores.

Using an electronic implementation, for example, a website, mobile app, or algorithm implemented into the electronic medical file of the patients, is more suitable to directly calculate the probabilities for each disease category. This electronic implementation is not only preferred for multinomial models, but also for dichotomous regression models.

4.1. Methodological considerations

The multinomial logistic regression model presented in this study was mainly developed for illustrative purposes. The data set provided the opportunity to apply multinomial regression modeling in a clinical situation where multiple conditions were considered simultaneously and allowed us to develop and present a multinomial nomogram to enhance presentation and understanding of such a model. However, the developed model cannot be applied in clinical practice yet. The model was developed in a relatively small data set for modeling four disease categories, and a rather large number of predictors were included into the model. Typically, external validation of a prediction model is needed before it can be considered for application in clinical practice. Testing the performance and amount of optimism of our multinomial model by internal validation was not performed. To our knowledge, no agreed statistical tool is available for testing optimism of a multinomial model in bootstrap samples. Further methodological work is needed in this area.

Multinomial models can be simplified by leaving some of the predictors out the model for a specific disease category, that is, setting these model coefficients to 0 and thereby reducing the number of tested model coefficients [24]. This method was not applied because the model was, as previously stated, developed for illustrative purposes.

Although the model was used for illustrative purposes, we did assess model performance with pairwise C statistics [22,23]. This method was identified as one of the preferred methods over the conventional C statistic (1-versus-rest) to assess the discriminative ability of the multinomial logistic regression model. The conventional C statistic is considered a suboptimal measure because the data have to be dichotomized into disease presence and absence to calculate the C statistic. By doing so, disease absence includes all other

disease categories, and the largest category will dominate the group. As a result, the C statistic is dominated by the discriminative ability between the largest group and the disease of interest and does not necessarily represent the true discriminative ability. For example, the conventional C statistic of HF is 0.79, which is largely dominated by the lack of discriminative ability with disease category Other (0.75, Fig. 1) and thereby masking the ability of the model to discriminate patients with HF from those with COPD (paired C statistic of 0.89). Disease classification was based on the category with the highest probability, which may not be the most informative way of classifying patients for clinical practice. Especially in patients for whom very similar predicted probabilities were derived from the model. For example, a patient with a BMI of 23.4 kg/m², a history of 8 pack years of smoking, NT-proBNP serum levels of 175 pg/mL, FEV1 of 64% of predicted, signs of fluid overload, and abnormal breathing sounds. According to the model, HF and COPD is the most likely disease category with a probability of 0.44, followed by COPD with a probability of 0.42. The treating physician of this patient has to decide whether he treats both HF and COPD (according to the highest probability) or COPD only, and either option may work for this patient. The classification tables should be seen as potential illustration of how to present such information.

In conclusion, a case study was used to illustrate the usefulness of a multinomial logistic regression prediction model in diagnostic research. The use of a multinomial nomogram enhances the applicability of multinomial regression models in daily practice.

Acknowledgments

The authors thank all participating GP practices (De Grebbe, Nieuw Rhenen, Kerkelanden, De Angstel, De Weegbree and Eedenburgh) and the personnel from Saltro laboratory (Utrecht, The Netherlands). The authors thank Maarten-Jan M Cramer, Jan-Willem J Lammers, Clemens G.K.M. Fauser, Mod Arif Nugroho Soenardi, Marcel Landman, and Harm Jan Huidekoper for their contribution to the panel diagnoses.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2015.10.016>.

References

- [1] Steyerberg EW. Clinical prediction models: a practical approach to development, validation and updating. New York: Springer; 2009.
- [2] Bender R, Grouven U. Using binary logistic regression models for ordinal data with non-proportional odds. J Clin Epidemiol 1998;51: 809–16.

- [3] Begg CB, Gray R. Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* 1984;71:11–8.
- [4] Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KG. Polytomous logistic regression analysis could be applied more often in diagnostic research. *J Clin Epidemiol* 2008;61:125–34.
- [5] Roukema J, van Loenhout RB, Steyerberg EW, Moons KG, Bleeker SE, Moll HA. Polytomous regression did not outperform dichotomous logistic regression in diagnosing serious bacterial infections in febrile children. *J Clin Epidemiol* 2008;61:135–41.
- [6] Wijesinha A, Begg CB, Funkenstein HH, McNeil BJ. Methodology for the differential diagnosis of a complex data set: a case study using data from routine CT scan examinations. *Med Decis Making* 1983;3:133–54.
- [7] Hosmer DW, Lemeshow S. Special topics, in applied logistic regression. *Applied logistic regression*. 2 ed. Hoboken: John Wiley & Sons, Inc; 2000.
- [8] Padeletti M, Jelic S, LeJemtel TH. Coexistent chronic obstructive pulmonary disease and heart failure in the elderly. *Int J Cardiol* 2008 10;125:209–15.
- [9] Rutten FH, Moons KG, Cramer MJ, Grobbee DE, Zuihthoff NP, Lammers JW, et al. Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. *BMJ* 2005;331:1379.
- [10] Rutten FH, Cramer MJ, Lammers JW, Grobbee DE, Hoes AW. Heart failure and chronic obstructive pulmonary disease: an ignored combination? *Eur J Heart Fail* 2006;8:706–11.
- [11] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
- [12] van Mourik Y, Moons KG, Bertens LC, Reitsma JB, Hoes AW, Rutten FH. Triage of frail elderly with reduced exercise tolerance in primary care (TREE). A clustered randomized diagnostic study. *BMC Public Health* 2012;12:385.
- [13] van Mourik Y, Bertens LC, Cramer MJ, Lammers JW, Reitsma JB, Moons KG, et al. Unrecognized heart failure and chronic obstructive pulmonary disease (COPD) in frail elderly detected through a near-home targeted screening strategy. *J Am Board Fam Med* 2014;27:811–21.
- [14] Broekhuizen BD, Sachs AP, Oostvogels R, Hoes AW, Verheij TJ, Moons KG. The diagnostic value of history and physical examination for COPD in suspected or known cases: a systematic review. *Fam Pract* 2009;26:260–8.
- [15] Broekhuizen BD, Sachs AP, Verheij TJ, Janssen KJ, Asma G, Lammers JW, et al. Accuracy of symptoms, signs, and C-reactive protein for early chronic obstructive pulmonary disease. *Br J Gen Pract* 2012;62:e632–8.
- [16] Kelder JC, Cramer MJ, van WJ, van TR, Mosterd A, Moons KG, et al. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. *Circulation* 2011;124:2865–73.
- [17] Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. *Health Technol Assess* 2009;13:1–207. [iii].
- [18] Oudejans I, Mosterd A, Bloemen JA, Valk MJ, van VE, Wielders JP, et al. Clinical evaluation of geriatric outpatients with suspected heart failure: value of symptoms, signs, and additional tests. *Eur J Heart Fail* 2011;13:518–27.
- [19] Roalfe AK, Mant J, Doust JA, Barton P, Cowie MR, Glasziou P, et al. Development and initial validation of a simple clinical decision tool to predict the presence of heart failure in primary care: the MICE (Male, Infarction, Crepitations, Edema) rule. *Eur J Heart Fail* 2012;14:1000–8.
- [20] Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
- [21] Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57.
- [22] Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learn* 2001;45:171–86.
- [23] van Calster B, Vergouwe Y, Looman CW, Van Belle V, Timmerman D, Steyerberg EW. Assessing the discriminative ability of risk models for more than two outcome categories. *Eur J Epidemiol* 2012;27:761–70.
- [24] Marshall RJ, Chisholm EM. Hypothesis testing in the polychotomous logistic model with an application to detecting gastrointestinal cancer. *Stat Med* 1985;4:337–44.