A generic nomogram for multinomial prediction models: theory and guidance for construction

Maarten van Smeden1*, Joris AH de Groot1, Stavros Nikolakopoulos1, Loes CM Bertens2, Karel GM Moons1 and Johannes B. Reitsma1

Abstract
Background: The use of multinomial logistic regression models is advocated for modeling the associations of covariates with three or more mutually exclusive outcome categories. As compared to a binary logistic regression analysis, the simultaneous modeling of multiple outcome categories using a multinomial model often better resembles the clinical setting, where a physician typically must distinguish between more than two possible diagnoses or outcome events for an individual patient (e.g., the differential diagnosis). A disadvantage of the multinomial logistic model is that the interpretation of its results is often complex. In particular, the calculation of predicted probabilities for the various outcomes requires a series of careful calculations. Nomograms are widely used in studies reporting binary logistic regression models to facilitate the interpretation of the results and allow the calculation of the predicted probability for individuals.

Methods and results: In this paper we outline an approach for deriving a generic nomogram for multinomial logistic regression models and an accompanying scoring chart that can further simplify the calculation of predicted multinomial probabilities. We illustrate the use of the nomogram and scoring chart and their interpretation using a clinical example.

Conclusions: The generic multinomial nomogram and scoring chart can be used irrespective of the number of outcome categories that are present.

Keywords: Prediction, Nomogram, Graphical presentation, Multinomial outcomes, Logistic model, Scoring chart

Background
The use of multinomial logistic regression modeling has been encouraged to study multiple unordered outcome categories (and possibly their combination) simultaneously [1, 2]. The multinomial logistic model can be considered to be an extension of the popular binary logistic regression model, which is often used in the presence of two mutually exclusive outcome categories. More specifically, multinomial logistic regression analysis can be viewed as a series of binary logistic regression analyses where one of the outcome categories is the reference category in each binary sub-model.

Researchers often collapse multiple outcome categories into a single binary (composite) outcome and use binary logistic regression models to analyze their data. Consequently, potentially important information about the different outcome categories is lost. It can also be argued that a binary outcome often does not accurately reflect clinical practice, where physicians commonly have to make decisions while considering more than two relevant choices. For instance, physicians often consider the presence of differential diagnoses (and prognoses) for an individual patient simultaneously [1].

One of the key reasons why researchers might refrain from multinomial logistic regression analysis is that the results from these models are more complex to interpret and more elaborate than results from a binary logistic regression analysis. In the multinomial context, regression coefficients are estimated for each binary sub-model.
reflecting the relation of covariates to one outcome category relative to the reference category. The number of estimated parameters quickly increases with additional outcome categories considered. The large amount of information from multinomial models can easily overwhelm researchers and clinicians. In addition, when using the multinomial logistic model for estimating probabilities for individual patients, the computation involved for the various outcomes requires a series of careful calculations.

Nomograms found many applications in the reporting of binary logistic regression models (for recent examples, see [3, 4]). A nomogram can not only improve insights of clinicians into the results of a logistic model, it can also be used to arrive at a predicted probability of outcome(s) of interest that is (are) tailored to the profile of an individual patient in a graphical manner. Nomograms can thus facilitate clinical decision making during clinical encounters. So far, nomograms have been used primarily for improving the reporting of models with only two outcome categories. We are aware of one recent paper focused on models [5] and noted that the focus of these papers was limited to constructing an nomogram for a limited number of outcome categories. We are aware of one recent paper focused on models [5] and noted that the focus of these papers was limited to constructing a nomogram for a limited number of outcome categories. However, the focus of these papers was limited to constructing a nomogram for a limited number of outcome categories.

In this manuscript, we present how to construct, interpret, and use a generic nomogram for a multinomial logistic prediction model. We will first specify the multinomial regression model and then present a general approach for deriving the nomogram and accompanying scoring chart for such models irrespective of the number of outcome categories that are present. We will illustrate the use of the nomogram and its interpretation using a clinical example on the risk of operative delivery [7].

**Multinomial logistic model**

Let $y_i$ denote the single observed outcome category of individual $i$. Assuming that this outcome is in one of $K$ categories (e.g., a disease among $K$ possible diseases), we may assume $Y_i$ to be a multinomial random variable with probabilities $\pi_{i1}, \ldots, \pi_{iK}$. Conditioned on $J$ observed covariate values in vector $x_i$, $x_i = \{x_{i1}, \ldots, x_{ij}, \ldots, x_{ij}\}$, the probability of observing category $k$ is denoted by $\pi_k(x_i)$. The multinomial logistic model where category $K$ is treated as the reference category, can then be defined as

$$\ln \frac{\pi_k(x_i)}{\pi_K(x_i)} = \ell p_k(x_i) = \alpha_k + \beta_k' x_i, \quad k = 1, \ldots, K - 1, \quad (1)$$

where $\alpha_k$ is an intercept term, $\beta_k = \{\beta_{k1}, \ldots, \beta_{kj}\}$ is a vector of regression coefficients and $\ell p_k(x_i)$ is one of $K - 1$ linear predictors for individual $i$. We can now define the probability of each possible category $k$ by

$$\pi_k(x_i) = \left\{ \begin{array}{ll}
\exp \{ \ell p_k(x_i) \} / \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] \\
1 / \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] & \text{if } k = 1, \ldots, K - 1 \\
1 / \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] & \text{if } k = K.
\end{array} \right. \quad (2)$$

**Methods: Constructing the multinomial nomogram**

The nomogram we suggest is a special case of a group of nomograms that are formally known as parallel scale nomograms. Doerfler [8] outlined the parallel scale nomogram that can be constructed if a particular value can be calculated from the sum of two functions. To use this approach for multinomial logistic models, we make use of a natural logarithm transformation applied to the elements of Eq. 2, such that,

$$\ln \pi_k(x_i) = \left\{ \begin{array}{ll}
\ell p_k(x_i) - \ln \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] \\
- \ln \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] & \text{if } k = 1, \ldots, K - 1 \\
- \ln \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right] & \text{if } k = K.
\end{array} \right. \quad (3)$$

To simplify notation, in the following we let $o_{ik} = \ln \pi_k(x_i)$, $l_{ik} = \ell p_k(x_i)$ and $s_i = - \ln \left[ 1 + \sum_{p=1}^{K-1} \exp \{ \ell p_p(x_i) \} \right]$. The parallel scale nomogram makes use of the relation: $o_{ik} = l_{ik} + s_i$. Each of these three elements in this relation corresponds to one of the three vertical axes of the nomogram. The axes are denoted by $L$ (left axis), $O$ (middle axis), and $S$ (right axis). Axis $L$ is a scaled function of linear predictor $k, m_1 l_{ik}$, where $m_1$ is the scaling factor. Axis O corresponds to the probability of observing category $k, m_2 o_{ik}$. Lastly, axis $S$ is a scaled function of the sum of exponentiated linear predictors, $m_3 s_i$. Axes $O$ and $S$ are on the natural log scale.

The nomogram is depicted in Fig. 1. Below we detail the four-step procedure to arrive at this nomogram. For further details about the construction of the parallel scale nomogram, we refer to Doerfler [8].

**Step 1: placing the outer axes ($L$ and $S$)**

To obtain an adequately sized nomogram, determine the desired common height ($h$) for the outer two axes ($L$ and $S$) and the horizontal distance between them ($d$). The two parallel axes are placed in the vertical direction. The values for $h$ and $d$ are assigned at the discretion of the researcher in a common metric (e.g., centimeters or inches). Larger values for $h$ and $d$ will allow for more precise reading of values.
Step 2: determine the scaling factors \( m_1, m_2, m_3 \)
For determining the scaling factors for the outer axes \((m_1\) and \(m_3\)), the relevant ranges of \( l_{ik}\) and \( s_i\) need to be considered. The limits of these ranges \((l_{\text{low}}, l_{\text{up}}, s_{\text{low}}, s_{\text{up}})\) may be determined by the range observed in the data set where the model was developed, e.g., \( l_{\text{low}} = \min(l_{ik}) \) and \( l_{\text{up}} = \max(l_{ik}) \). These limits define the corresponding limits of the axes. Once these ranges are chosen, the scaling factors \( m_1 \) and \( m_3 \) are computed by
\[
m_1 = \frac{h}{l_{\text{up}} - l_{\text{low}}} \quad \text{and} \quad m_3 = \frac{h}{s_{\text{up}} - s_{\text{low}}}.
\]
The remaining scaling factor is given by \( m_2 = m_1m_3 / (m_1 + m_3) \).

Step 3: placing the middle axis (O)
The O axis is placed parallel to the outer axes. The horizontal distance between the axes L and O is given by \( d_{LO} = d - d/(m_1/m_3 + 1) \).

Step 4: placing tick marks and labels
For the outer axes L and S, two sequences of values for the tick marks and corresponding labels are defined: \( l_T = (l_{\text{low}}, l_1, \ldots, l_{\text{up}}) \) and \( s_T = (s_{\text{low}}, s_1, \ldots, s_{\text{up}}) \). Tick mark \( t \) on the axis should be placed relative to the lower end of that axis at a distance of \( m_1 \times (l_t - l_{\text{low}}) \) for axis L and \( m_3 \times (s_t - l_{\text{low}}) \) for axis S. For axis O, we define: \( o_{\text{low}} = \exp(l_{\text{low}} + s_{\text{low}}) \). Because the axis O represents a log transformed probability scale, first a sequence of arithmetic probabilities \( o_T^* \) is defined with values between \( \exp(o_{\text{low}}) \) and 1. Then, tick mark \( t \) for this sequence may be placed at \( m_2 \times (\ln(o_T^*) - f(o_{\text{low}})) \) labeled by \( o_T^* \). Axis S is labeled by \( \sum_{p=1}^{K-1} \exp(l_p(x_i)) \).

Methods: Constructing the scoring chart
The use of the nomogram by health professionals can be improved by additionally presenting a scoring chart. This scoring chart provides a graphical approach to arriving at the values for the two outer axes L and S of the nomogram for any relevant combination of values on the covariates \((x_i)\). For brevity, in this section we only consider the case of a multinomial logistic regression model with first order main effects. The scoring chart (and nomogram) can be extended to accommodate situations where higher order and interaction effects are present.

To make the scoring chart user-friendly, the individual effects of covariate \( j \), \( \hat{\beta}_{jk} x_{ij} \), that make up the linear predictor \( k \) are rescaled to a “standardized” score. The sum over these individual effects together with a baseline score make up a “standardized” total score. This total score is a linear transformation of \( l_{ik} \). To facilitate the applicability of this standardized total score approach to the nomogram, the scaling of axis L should be adjusted accordingly. Below we detail the three-step procedure to arrive at the scoring system that makes up the scoring chart.
Step 5: standardized covariate effects: points

The estimated multinomial logistic regression coefficients, $\hat{\beta}_{jk}$, are rescaled relative to the largest (conditional) covariate effect on a scale that has a minimum of 0 and a maximum of 100. First, the relevant ranges for each of the covariate variables are considered. Let the boundaries of these relevant ranges be denoted: $x_{j}^{\text{low}}$ and $x_{j}^{\text{up}}$. The rescaling factor and rescaled coefficients are then computed by $r = 100/\max_{j,k}(|\hat{\beta}_{jk} x_{j}^{\text{up}}|, \hat{\beta}_{jk} x_{j}^{\text{low}}|)$ and $\hat{\beta}_{jk}^{*} = r \times \hat{\beta}_{jk}$. The covariate effects are “standardized” by $\text{Points}_{jk}(x_{ij}) = \hat{\beta}_{jk}^{*} x_{ij} - \min(\hat{\beta}_{jk}^{*} x_{j}^{\text{up}}, \hat{\beta}_{jk}^{*} x_{j}^{\text{low}})$.

Step 6: standardized total effect: total points

A baseline score for each category (except the reference category) is defined that takes into account the restricted range of these relevant ranges be denoted: $x_{j}^{\text{low}}$ and $x_{j}^{\text{up}}$. The baseline score is computed by $b_{lk} = r \times \hat{\alpha}_{k} + \sum_{j} \min(\hat{\beta}_{jk}^{*} x_{j}^{\text{up}}, \hat{\beta}_{jk}^{*} x_{j}^{\text{low}})$. To also obtain a “standardized” baseline score such that the minimum rescaled baseline score is zero, we subtract the minimum baseline score, $b_{lk}^{*} = b_{lk} - \min_{k}(b_{lk})$. The standardized total effect for category $k$ given covariate values is then given by $\text{Total}_{k} = b_{lk}^{*} + \sum_{j} \text{Points}_{jk}(x_{ij})$. Notice that $l_{ik} = \text{lp}_{k}(x_{i}) = (\text{Total}_{k} + \min_{k}(b_{lk}))/r$.

Step 7: connecting the standardized total effects to the S axis

A horizontal axes representing $\text{Total}_{k}$ is placed near the lower end of the scoring chart. Another parallel horizontal axis is placed: the values on this axis are related to the former axis by $\exp(\text{Total}_{k} + \min_{k}(b_{lk}))$. Taking the sum over the values that can be read off from the axis for all categories (except the reference category) is all the information necessary for determining the value on the $S$ axis.

Result and discussion

Empirical example: predicting the risk of operative delivery

To illustrate the suggested scoring chart and nomogram for the reporting of multinomial logistic models, we use a previously published model on predicting the risk of operative delivery [7]. We detail their use by considering a specific hypothetical subject (described below). The prediction model used in this illustration was developed using data from a randomized clinical trial conducted in the Netherlands [9].

In brief, the multinomial prediction model was developed in 5667 laboring women with high-risk vertex (i.e., babies in a normal position in the uterus) singleton pregnancies beyond 36 weeks of gestation that met the inclusion criteria of the randomized clinical trial. Based on the combination of the intervention (i.e., instrumental vaginal delivery (IVD) or caesarean section (CS)) and the indication for the intervention (i.e., fetal distress (FD) or failure to progress (FTP)), women were assigned to one of five distinctive outcome categories: spontaneous vaginal delivery (reference category); instrumental vaginal delivery due to suspected fetal distress (IVD-FD); caesarean section due to suspected fetal distress (CS-FD); instrumental vaginal delivery due to failure to progress (IVD-FTP); or caesarean section due to failure to progress (CS-FTP). The multinomial regression model included the antepartum variables: maternal age, parity, gestational age, maternal diabetes mellitus, previous caesarean delivery, fetal gender, maternal hypertensive disorder, suspected intrauterine growth restriction, and antepartum estimated fetal weight. An antepartum prediction model was developed using this set of variables (i.e., model 1 in Schuit et al. 2012; see Table 1). For more details on the various outcome categories and candidate predictors, we refer to the original publication [7].

Table 1 Multivariable associations for multinomial antepartum prediction model, predicting the risk of operative delivery

|                      | IVD-FD vs spont. | CS-FD vs spont. | IVD-FTP vs spont. | CS-FTP vs spont. |
|----------------------|------------------|-----------------|-------------------|------------------|
| $\beta_{k}$ OR(95% CI) | $\beta_{k}$ OR(95% CI) | $\beta_{k}$ OR(95% CI) | $\beta_{k}$ OR(95% CI) |
| Intercept             | −13.1            | −15.6           | −11.1             | −15.4            |
| Maternal age, years   | 0.029 (1.01, 1.05)| 0.052 (1.01, 1.09)| 0.054 (1.03, 1.08)| 0.056 (1.04, 1.08)|
| Gestational age, weeks| 0.26 (1.18, 1.41)| 0.32 (1.22, 1.56)| 0.038 (0.95, 1.13)| 0.13 (1.05, 1.24)|
| Nulliparous           | 2.05 (5.26, 11.5)| 1.13 (3.09, 4.55)| 3.39 (29.7, 51.1)| 2.65 (14.1, 20.3)|
| Previous caesarean delivery | 1.77 (3.70, 9.32)| 1.06 (2.88, 4.76)| 2.39 (10.9, 20.1)| 2.23 (9.34, 16.1)|
| Neonatal female gender| −0.19 (0.67, 1.03)| 0.05 (0.61, 0.83)| −0.25 (0.78, 1.06)| −0.013 (0.99, 1.20)|
| Birthweight, 100-g increments | −0.059 (0.94, 0.97)| −0.079 (0.92, 0.96)| 0.083 (1.09, 1.11)| 0.12 (1.12, 1.15)|
| Maternal diabetes mellitus | 0.32 (1.37, 2.91)| 0.99 (2.69, 5.60)| −0.24 (0.79, 1.76)| 0.87 (2.38, 3.95)|
Consider the following hypothetical subject: a nulliparous diabetic subject with a maternal age of 32 years, a gestational age of 40 weeks, expecting a boy with an estimated birth weight of 3540 g. Using the results presented in Table 1 and a series of calculations (that follow from Eqs. 1 and 2), one may calculate the predicted probabilities for the hypothetical case study for each of the five outcomes as 0.096 (IVD-FD), 0.051 (CS-FD), 0.069 (IVD-FTP), 0.268 (CS-FTP), and 0.516 (spontaneous delivery). To facilitate the calculation of the predicted probabilities, we constructed a scoring chart (Fig. 2) and nomogram (Fig. 2) using the reported regression coefficients by Schuit et al. [7]. For illustrative purposes we have added some information specific to the case study in Figs. 2 and 2. The scoring chart and nomogram without the case study information can be found in Webfigure A and Webfigure B, respectively.

Using the scoring chart (Fig. 2), each subject characteristic can be converted into a score for a particular outcome category (except the reference category). We have illustrated this procedure by drawing vertical dashed lines between the axes only for the subject characteristic

---

**Fig. 2** Scoring chart—hypothetical case study based on multinomial prediction model in Schuit et al. Case description: Maternal age: 32 years; gestational age: 40 weeks; nulliparous; birth weight: 3540 g, maternal diabetes. Abbreviations: instrumental vaginal delivery (IVD), caesarean section (CS), fetal distress (FD), failure to progress (FTP)
“maternal age” (32 years) to the Points axis (upper end scoring chart) in the figure. We leave it to the interested reader to draw the lines for the remaining characteristics. By adding the points of the separate predictors to the baseline points for each of the chosen outcomes, one obtains the total points for this particular outcome. Each of these total points should be marked at the line indicated by Total Points (lower end scoring chart). For the considered subject, the total points are 159 (IVD-FD), 140 (CS-FD), 149 (IVD-FTP), and 189 (CS-FTP). Further, by drawing vertical lines from the marked points on Total Points axis to the Exp$(l_{i,k})$ axis (illustrated using dashed lines), one can read-off the exponentiated linear predictor for each of the outcome categories. This information is then put in the boxes right under the Exp$(l_{i,k})$ axis. Finally, by taking the sum over these values (registered under box “S”) one obtains all information needed to use the nomogram (Fig. 2).

To use the nomogram (Fig. 2), one first marks on the Total Points axis (left axis) the total points for each outcome category except the reference category, as derived from the scoring chart. One than marks the value registered under box “S” in the scoring chart on the $S$ axis of (Fig. 2). The points marked on the Total points axis and point on the $S$ axis are then connected using straight lines. From the middle “$O$” axis, one can now read-off the predicted probabilities for each of the four outcomes for this particular case. Finally, by subtracting these four probabilities from one, the probability of the reference category is obtained, corresponding to the probability of a “spontaneous delivery”.

**Conclusion**

Nomograms are not likely to be used very often in contemporary clinical practice for calculating probabilities for individual patients given the considerable burden placed on the user and also because its accuracy is limited by the precision with which physical markings can be drawn, reproduced, viewed, and aligned. For this purpose, electronic implementation of a prediction model by means of a website calculator, mobile app, or an algorithm implemented into the electronic medical file of the patients to directly calculate the relevant probabilities may be considered. However, we still think that nomograms remain quite useful, even today, as they do offer great visual insight a (multinomial) prediction model, which is often perceived as a black-box (for a further discussion on the usefulness of nomograms we refer to [10, 11]). Therefore, we do expect that our general approach to construct and report scoring charts and nomograms for multinomial logistic regression models will facilitate the interpretation and use of such models to a wider audience.
Our approach is flexible and generalizable and can be used irrespective of the number of outcome categories and types of covariates present. R-code for developing the nomogram is available from http://mvansmeden.net/software/multinomial-nomogram.

Abbreviations
CS: Caesarean section; FD: Fetal distress; FTP: Failure to progress; IVD: Instrumental vaginal delivery

Acknowledgements
Not applicable.

Funding
We gratefully acknowledge financial contribution from the Netherlands Organisation for Scientific Research (project 918.10.615).

Availability of data and materials
Not applicable.

Authors’ contributions
MvS, JdG, JR, LB, and KM drafted the concept for the current paper. MvS wrote the initial version of the paper. SN contributed to the statistical notation. All authors commented on drafts of the article and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
1 Julian Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, Netherlands. 2 Department of Obstetrics and Gynaecology, Erasmus MC, Rotterdam, Netherlands.

Received: 9 September 2016 Accepted: 13 January 2017

Published online: 10 April 2017

References
1. Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KGM. Polytomous logistic regression analysis could be applied more often in diagnostic research. J Clin Epidemiol. 2008;61(2):125–34.
2. Begg CB, Gray R. Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions. Biometrika. 1984;71(1):11.
3. Ogura K, Fujiwara T, Yashunaga H, Matsui H, Jeon DG, Cho WH, Hiraga H, Ishii T, Yonemoto T, Kamoda H, et al. Development and external validation of nomograms predicting distant metastases and overall survival after neoadjuvant chemotherapy and surgery for patients with nonmetastatic osteosarcoma: A multi-institutional study. Cancer. 2015;121(21):3844–52.
4. Aviram G, Shmueli H, Adam SZ, Bendet A, Ziv-Baran T, Steinvil A, Berliner AS, Nesher N, Ben-Gal Y, Topilsky Y. Pulmonary Hypertension: A Nomogram Based on CT Pulmonary Angiographic Data for Prediction in Patients without Pulmonary Embolism. Radiology. 2015;277(1):236–46.
5. Ardioino I, Lanzoni M, Marano G, Boracchi P, Sagnin E, Gian Stefani A, Piscaglia F, Biganzoli EM. Widien Nomogram for multinomial logistic regression: an application to staging liver fibrosis in chronic hepatitis C patients. Stat Methods Med Res. 2014. [Epub].
6. Bertens LCM, Moons KGM, Rutten FH, van Mourik Y, Hoes AW, Reitsma JB. A nomogram was developed to enhance the use of multinomial logistic regression modeling in diagnostic research. J Clin Epidemiol. 2016;71:51–71.
7. Schuit E, Kwee A, Westerhuis M, Van Dessell H, Graziosi G, Van Lith J, Nijhuis J, Oei S, Oosterbaan H, Schuurman N, et al. A clinical prediction model to assess the risk of operative delivery. BJOG. 2012;119(8):915–23.
8. Doerfler R. On Jargon-The Lost Art of Nomography. UMAP J. 2009;30(4):457.
9. Westerhuis MEMH, Visser GHA, Moons KGM, van Beek E, Benders MJ, Bijvoet SM, van Dessel HJHM, Drogtrop AP, van Geijn HP, Graziosi GC, et al. Cardiotocography Plus ST Analysis of Fetal Electrocardiogram Compared With Cardiotocography Only for Intrapartum Monitoring. Obstet Gynecol. 2010;116(6):173–80.
10. Marshall RJ. Are regression “nomograms” useful? J Clin Epidemiol. 2016;78:4–6.
11. Reitsma JB, van Smeden M, de Groot JAH, Bertens LCM, Moons KGM. Response to the commentary on “A nomogram was developed to enhance the use of multinomial logistic regression modelling in diagnostic research”. J Clin Epidemiol. 2016;78:7–9.