PSYCHOMETRIKA—VOL. 56, NO. 4, 589–600
DECEMBER 1991

A GENERALIZED RASCH MODEL FOR MANIFEST PREDICTORS

AEILKO H. ZWINDERMAN

DEPARTMENT OF MEDICAL STATISTICS
UNIVERSITY OF LEIDEN

A logistic regression model is suggested for estimating the relation between a set of manifest predictors and a latent trait assumed to be measured by a set of k dichotomous items. Usually the estimated subject parameters of latent trait models are biased, especially for short tests. Therefore, the relation between a latent trait and a set of predictors should not be estimated with a regression model in which the estimated subject parameters are used as a dependent variable. Direct estimation of the relation between the latent trait and one or more independent variables is suggested instead. Estimation methods and test statistics for the Rasch model are discussed and the model is illustrated with simulated and empirical data.

Key words: latent trait, Rasch model, logistic regression, conditional marginal likelihood.

Introduction

Attempts to relate (psychological) scales to other variables are often based on linear models. In the classical approach to test theory (see Lord & Novick, 1968), the weighted or unweighted test scores are typically linearly related to m independent variables (called predictors in the sequel):

$$Y_v = \beta_0 + \beta_1 x_{v1} + \cdots + \beta_m x_{vm} + \varepsilon_v. \tag{1}$$

Here, Y_v is the test score of individual v, x_{vj} is the observed value of predictor j for individual v, β_j is the corresponding regression parameter, and ε_v is the error term.

In contrast to classical test theory, latent trait theory (see Fischer, 1974; Lord, 1980) uses the item responses to estimate latent variables. In the dichotomous Rasch model, the conditional probability of a positive response of subject v on item i ($Y_{vi} =$ 1) is modeled as a function of the latent (ability) parameter of $v(\theta_v)$ and the latent (difficulty) parameter of $i(\alpha_i)$ with the familiar logistic function:

$$p(Y_{vi} = 1|\theta_v) = \frac{\exp(\theta_v - \alpha_i)}{1 + \exp(\theta_v - \alpha_i)}. \tag{2}$$

The item parameters $\boldsymbol{\alpha}$ can be estimated consistently either with conditional maximum likelihood (see Andersen, 1970, for the Rasch model only) or with marginal maximum likelihood (see Bock & Aitkin, 1981). But for the estimation of the relation between the latent trait and one or more predictors, the subject parameters $\boldsymbol{\theta}$ are required as a dependent variable. The subject parameters can be estimated with maximum likelihood or with the mean or mode of the posterior distribution when using marginal maximum likelihood estimation. However, the estimates of the subject parameters ($\hat{\boldsymbol{\theta}}$) are biased and inconsistent (see Goldstein, 1980; Lord, 1984). The inconsistency of $\hat{\boldsymbol{\theta}}$ makes it problematic for use in regression models.

In this paper a method is suggested to estimate the relation between the latent trait

Requests for reprints should be sent to Aeilko H. Zwinderman, Department of Medical Statistics, University of Leiden, P.O. Box 9512, 2300 RA Leiden, THE NETHERLANDS.

0033-3123/91/1200-89092$00.75/0

and one or more predictors directly without estimating the subject parameters. The method is developed for the Rasch model, but can be generalized easily to the two- and three-parameter logistic latent trait models (see Lord & Novick, 1968). In the following section we will develop the model and consider methods for estimation and testing.

The Logistic Regression Model

Model

Consider the population of individuals concerning the latent variable θ with density function, $g(\theta)$. A sample of N individuals responds to a questionnaire containing k items that are scored dichotomously. The probability that individual $v(v = 1, \ldots, N)$ responds positively to item $i(i = 1, \ldots, k)$ is modeled with the Rasch model as specified in (2). Now consider the (linear) model for θ:

$$\theta_v = \boldsymbol{\beta}'\mathbf{x}_v + \varepsilon_v, \tag{3}$$

where $\mathbf{x}_v$ is a vector of length m consisting of the observations for individual v on m predictors, $\boldsymbol{\beta}$ is the vector of the unknown regression parameters, and ε_v is the usual error term. Equation (3) may include interaction or higher-order terms. Substituting (3) in (2) gives

$$p(Y_{vi} = 1|\varepsilon_v, \mathbf{x}_v) = \frac{\exp(\beta_0 + \beta_1 x_{v1} + \cdots + \beta_m x_{vm} + \varepsilon_v - \alpha_i)}{1 + \exp(\beta_0 + \beta_1 x_{v1} + \cdots + \beta_m x_{vm} + \varepsilon_v - \alpha_i)} \equiv h_{vi}. \tag{4}$$

Let $x_1, \ldots, x_m$ be fixed, and ε_v independent, identically distributed normal random variables with mean zero and standard deviation σ, and density function $\phi(\varepsilon|\sigma)$. This defines (4) as a logistic regression model with fixed effects for the predictors, and a random component. For models like this, see Breslow and Day (1980).

The joint likelihood of the item response vector $\mathbf{Y}_v = (Y_{v1}, \ldots, Y_{vk})$ and ε_v given $\mathbf{x}_v$ is

$$p(\mathbf{Y}_v, \varepsilon_v|\mathbf{x}_v) = p(\mathbf{Y}_v|\varepsilon_v, \mathbf{x}_v)\phi(\varepsilon|\sigma) = \prod_{i=1}^{k} h_{vi}^{y_{vi}}(1 - h_{vi})^{1-y_{vi}}\phi(\varepsilon|\sigma),$$

where y_{vi} is the realization of Y_{vi}. The marginal likelihood of $\mathbf{Y}_v$ given $\mathbf{x}_v$ is

$$p(\mathbf{Y}_v|\mathbf{x}_v) = \int_{-\infty}^{\infty} p(\mathbf{Y}_v|\varepsilon_v, \mathbf{x}_v)\phi(\varepsilon|\sigma)\, d\varepsilon,$$

and the conditional marginal likelihood of all item responses, Y, given X, $L_m(Y|X, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$, is

$$L_m(Y|X, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma) = \prod_{v=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{k} h_{vi}^{y_{vi}}(1 - h_{vi})^{1-y_{vi}}\phi(\varepsilon|\sigma)\, d\varepsilon. \tag{5}$$

The conditional marginal maximum likelihood estimators $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma})$ maximizing (5) are consistent and asymptotically normally distributed (Bock & Aitkin, 1981). The number of parameters is finite, and therefore, standard ML asymptotics apply.

The model specified in (4) cannot be estimated uniquely. A restriction is needed both for $\boldsymbol{\alpha}$ and the distribution of ε. We norm the item parameters in such a way that the mean of the estimates of $\boldsymbol{\alpha}$ is zero, and the mean of ε is fixed to zero. Alternatively, the mean of ε might be free, but then the intercept parameter β_0 must be fixed to zero. If all item parameters are allowed to vary, both the intercept and the mean of ε must be fixed to zero.

Sampling Designs

The likelihood in (5) is derived from the conditional distribution of the item responses given the predictors: $p(Y|X)$. This is the appropriate likelihood in case of experimental studies, where the predictors have been fixed a priori by the experimenter. This kind of sampling scheme is called conditional sampling (see J. A. Anderson, 1972). It arises, for instance, if the experimenter is interested in the differences between males and females or between ethnic groups with respect to the latent variable θ.

With observational studies, the predictors are not fixed a priori, but are sampled from some population with (multivariate) density function $G(X)$. In this case, the joint likelihood of $\mathbf{Y}_v$, $\mathbf{X}_v$, and ε_v is given by

$$p(\mathbf{Y}_v,\ \varepsilon_v,\ \mathbf{X}_v) = p(\mathbf{Y}_v,\ \varepsilon_v|\mathbf{x}_v)G(\mathbf{X}_v) = p(\mathbf{Y}_v|\varepsilon_v,\ \mathbf{x}_v)\phi(\varepsilon_v|\sigma)G(\mathbf{X}_v),$$

and the marginal likelihood is now given by

$$L_m(Y,\ X|\boldsymbol{\alpha},\ \boldsymbol{\beta},\ \sigma) = \left[\prod_{v=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{k} h_{vi}^{y_{vi}}(1 - h_{vi})^{1-y_{vi}}\phi(\varepsilon|\sigma)\ d\varepsilon\right]\left[\prod_{v=1}^{N} G(\mathbf{X}_v)\right]. \tag{6}$$

The second term of (6) ($\Pi_v G(\mathbf{X}_v)$) is uninformative with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ, because $G(X)$ is not a function of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, or σ. Therefore we need to optimize only the first term of (6) to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ. This part of the likelihood is equal to the conditional marginal likelihood as specified in (5). Hence, the ML estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ in case of observational studies are equal to the ML estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ with conditional sampling. Anderson (1972) and van Houwelingen and le Cessie (1988) call this kind of sampling, mixture sampling. The asymptotic results of conditional sampling also apply to the estimators derived under mixture sampling (Prentice & Pyke, 1979).

The third case where a logistic regression model arises, is the case of separate sampling. This sampling design generalizes from case-control studies (see Breslow & Day, 1980). A separate sampling design arises when the item response vector $\mathbf{y}_v$ is fixed. Prentice and Pyke (1979), Breslow and Day, and van Houwelingen and le Cessie (1988) showed that the estimators derived under conditional sampling can also be seen as the maximum likelihood estimators under separate sampling. But matters are more delicate. With k items, there are 2^k different response patterns. Separate sampling means that for every response pattern $\mathbf{y}_r$, a sample of n_r is taken from the conditional distribution of $(X,\ \varepsilon)$ given $\mathbf{Y} = \mathbf{y}_r$. The likelihood of the observations is given by

$$L(X) = \prod_{r=1}^{2^k} \prod_{v=1}^{n_r} f_r(\mathbf{X}_v,\ \varepsilon_v|\mathbf{y}_r),$$

where $f_r(\mathbf{X}_v,\ \varepsilon_v|\mathbf{y}_r)$ is the conditional distribution of $(X,\ \varepsilon)$ given $Y = \mathbf{y}_r$. Let

$$p_r = \frac{n_r}{n}, \quad \text{with} \quad n = \sum_{r=1}^{2^k} n_r,$$

$$f(X, \varepsilon) = \sum_{r=1}^{2^k} p_r f_r(X, \varepsilon|\mathbf{y}_r), \quad \text{and}$$

$$g_r(X, \varepsilon) = \frac{p_r f_r(X, \varepsilon|\mathbf{y}_r)}{f(X, \varepsilon)}.$$

The logistic model assumption is equivalent to

$$g_r(X, \varepsilon) = \prod_{i=1}^{k} g_{ri}(X, \varepsilon),$$

$$\text{logit}\ [g_{ri}(X, \varepsilon)] = \boldsymbol{\beta}'\mathbf{X} + \varepsilon - \alpha_i$$

(van Houwelingen & le Cessie, 1988, pp. 220–221). Furthermore, assuming X and ε to be independent, $f(X, \varepsilon)$ can be written as the product $m(X)\phi(\varepsilon|\sigma)$, where $m(X)$ is a (multivariate) density function of the predictors X. The likelihood $L(X)$ can now be rewritten as

$$L(X) = \left[\prod_{v=1}^{N} m(\mathbf{X}_v)\right] \times \left[\prod_{v=1}^{N}\prod_{i=1}^{k} h_{vi}^{y_{vi}}(1 - h_{vi})^{1-y_{vi}}\phi(\varepsilon|\sigma)\right] \times \left[\prod_{r=1}^{2^k} p_r^{-n_r}\right]$$
$$= L_1 \times L_2 \times \text{constant},$$

where h_{vi} is defined as in (4). The maximum likelihood principle means that $L(X)$ is maximized with respect to $m(X)$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ, integrating out ε. Assume for simplicity that X has a discrete distribution with generic value z. The unrestricted maximum likelihood estimator of $m(z)$, $\hat{m}(z)$, is the fraction of subjects with $x = z$. The unrestricted maximum of L_2, integrating out ε, is given by the estimator $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma})$ as derived under conditional sampling (see Prentice and Pyke, 1979, for more details).

Separate sampling designs are important in medical research, but it will seldom be appropriate in applied psychometrics. An example of separate sampling in psychometrics is when one is studying the characteristics of (groups of) individuals with some specified response pattern.

Estimation of the Model Parameters

We developed an EM algorithm to estimate the parameters of the model specified in (4), which is comparable to the one developed by Rigdon and Tsutakawa (1983). We maximize the expected value of the joint log-density function of $(Y, \varepsilon|X)$, $m(Y, \varepsilon|X) = \log(P(Y, \varepsilon|X))$, given starting values of the model parameters, $\hat{\boldsymbol{\alpha}}_{(0)}$, $\hat{\boldsymbol{\beta}}_{(0)}$, $\hat{\sigma}_{(0)}$,

$$F = E[m(Y, \varepsilon|X)|\hat{\boldsymbol{\alpha}}_{(0)}, \hat{\boldsymbol{\beta}}_{(0)}, \hat{\sigma}_{(0)}]$$
$$= \sum_{v=1}^{N}\sum_{i=1}^{k}\int_{-\infty}^{\infty} (y_{vi}\log(h_{vi}) + (1 - y_{vi})\log(1 - h_{vi}))p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)\, d\varepsilon$$
$$+ \sum_{v=1}^{N}\int_{-\infty}^{\infty}\log\left((2\pi\sigma^2)^{1/2}\exp\left(-\frac{1}{2}\frac{\varepsilon^2}{\sigma^2}\right)\right)p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)\, d\varepsilon, \tag{7}$$

where $p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)$ is the posterior density function of ε given $(\mathbf{x}_v, \mathbf{y}_v)$ and $(\hat{\boldsymbol{\alpha}}_{(0)}, \hat{\boldsymbol{\beta}}_{(0)}, \hat{\sigma}_{(0)})$. The posterior density function is proportional to

$$p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v) \propto \frac{\exp(y_{v.}\ \varepsilon_v) \exp\left(-\frac{1}{2}\frac{\varepsilon_v^2}{\sigma_{(0)}^2}\right)}{\prod_{i=1}^{k} 1 + \exp(\hat{\boldsymbol{\beta}}'_{(0)}\mathbf{x}_v + \varepsilon_v - \hat{\alpha}_{i(0)})},$$

where $y_{v.}$ is the raw score of subject v.

The algorithm is iterative with two steps. In the E-step, (7) is maximized with respect to σ. Therefore, the partial derivative of F with respect to σ^2 is equated to zero, yielding the following equation to solve for σ^2:

$$\sigma^2 = \frac{1}{N} \sum_{v=1}^{N} \int_{-\infty}^{\infty} \varepsilon^2 p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)\, d\varepsilon. \tag{8}$$

In the M-step (7) is maximized with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, simultaneously. Therefore, the partial derivatives of F with respect to $\alpha_i (i = 1, \ldots, k)$ and $\beta_j (j = 1, \ldots, m)$ are equated to zero, yielding the following two sets of equations to solve for α_i and β_j, respectively:

$$y_{.i} = \sum_{v=1}^{N} \int_{-\infty}^{\infty} \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_v + \varepsilon - \alpha_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_v + \varepsilon - \alpha_i)} p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)\, d\varepsilon, \tag{9}$$

and

$$\sum_{v=1}^{N} x_{vj} y_{v.} = \sum_{v=1}^{N} x_{vj} \int_{-\infty}^{\infty} \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_v + \varepsilon - \alpha_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_v + \varepsilon - \alpha_i)} p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)\, d\varepsilon, \tag{10}$$

where $y_{.i}$ is the item score of item i, and $y_{v.}$ is the raw score of subject v. The two sets of equations can be solved with any efficient optimization algorithm. The integrals can be evaluated with Gauss-Hermite quadrature (see Abramowitz & Stegun, 1964). Standard errors of the estimated model parameters can be obtained from the matrix of second order partial derivatives of (5) evaluated at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma})$. The partial second order derivatives of (5) are derived in the Appendix.

For the Rasch model, the algorithm can be simplified by using the conditional maximum likelihood estimates of $\boldsymbol{\alpha}$ as fixed constants in (5), (7), (8), and (10). This means that (9) is eliminated from the M-step as specified above. Such a two-stage procedure was also suggested by Andersen and Madsen (1977) for the estimation of the population distribution of θ. Although some efficiency in estimating $\boldsymbol{\alpha}$ is lost, the reduction in the computational effort is enormous.

Testing the Model

The model specified in (4) is a generalization of the Rasch model. Hence, the assumptions that have been made for the model as specified in (4) are equivalent to those of the Rasch model (see Fischer, 1974):

i. Monotonicity: The probability of a positive response is a monotone increasing function of $\boldsymbol{\beta}'\mathbf{x} + \varepsilon$.

ii. Conditional independence: Given fixed level of $\boldsymbol{\beta}'\mathbf{x} + \varepsilon$, the responses to the items are independent of other items.

iii. Unidimensionality: The error component ε is independent of the item parameters.

iv. The regression parameters $\boldsymbol{\beta}$ are independent of the item parameters.

The assumptions (i) and (iv) together are equivalent to the assumption of the simple Rasch model that the item characteristic curves are parallel. The assumptions (i) to (iv) can be tested with statistics that were developed to test the Rasch model (see Andersen, 1973; Glas, 1988, 1989).

Two additional assumptions are:

v. The error component is normally distributed.

vi. The regression between θ and $\boldsymbol{\beta}'\mathbf{x} + \varepsilon$ is linear.

As yet, we have not developed a global test of the model as specified in (4) that would be sensitive and informative for violation of assumptions (v) and (vi).

Hypotheses with respect to $\boldsymbol{\beta}$ can be tested with likelihood ratio statistics, Wald, or scoring type statistics (see Rao, 1973, pp. 415–420). Wald and scoring statistics are useful for stepwise selection of predictors. The Wald statistic needs only the full model (with all predictors included) to be estimated, and is therefore useful for backward selection. The scoring statistic needs only the restricted model (with a selected number of predictors included) to be estimated, and is therefore useful in forward selection. Several other test statistics and heuristic fit procedures can be generalized from the logistic regression model, for example, log-odds plot, and log-log plot per item (see BMDP; Dixon, 1985).

Simulations

Some simulations were performed to compare our method for estimating the regression of θ on x to the method of estimating the regression with a linear model on the estimated subject parameters

$$\hat{\theta}_v = \boldsymbol{\beta}'\mathbf{x}_v + \varepsilon_v. \tag{11}$$

We expect that our method described in the previous section performs better in estimating $\boldsymbol{\beta}$ than using (11) when the number of items is small, because in that case the distribution of $\hat{\theta}$ is very discrete and the variance and bias of $\hat{\theta}$ are large. Therefore, we simulated item responses for varying number of items (5, 10, 20, 30). The item difficulties were equal to zero, or spaced over the range $(-2, 2)$. Samples of 100 or 1000 individuals were taken.

Two predictors and an error term, (X_1, X_2, ε), were sampled from the multivariate normal distribution with mean vector zero, and diagonal covariance matrix. In all simulations, the subject abilities were calculated as $\theta_v = -1/2\, x_{v1} + 1/2\, x_{v2} + \varepsilon_v$. The variances of x_1 and x_2 were equal to one, and the variance of ε was 0.50. Consequently, the variance of θ was equal to one. Item responses were simulated by comparing the item response probabilities according to (2) to random numbers sampled from the uniform distribution with domain (0, 1). If the probability was larger than the random number, the item response was one, and zero otherwise.

The item parameters were estimated with conditional maximum likelihood (CML)

TABLE 1

Estimates of Regression Parameters According to (4) and (11). The Standard Errors (se) are Given in Parentheses

k	$\hat{\beta}_1$ se	$\tilde{\beta}_1$ se	$\hat{\beta}_2$ se	$\tilde{\beta}_2$ se
		N=1000, $\alpha \in (-2,2)$		
30	-.42(.014)	-.45(.037)	.52(.014)	.55(.036)
20	-.49(.017)	-.48(.037)	.46(.017)	.48(.037)
10	-.51(.024)	-.42(.042)	.47(.025)	.40(.043)
5	-.40(.035)	-.28(.040)	.56(.035)	.41(.040)
		N=100, $\alpha \in (-2,2)$		
30	-.43(.047)	-.50(.118)	.57(.047)	.60(.114)
20	-.47(.047)	-.56(.099)	.42(.054)	.45(.114)
10	-.49(.073)	-.38(.140)	.39(.080)	.32(.149)
5	-.29(.101)	-.20(.112)	.51(.108)	.46(.123)
		N=1000, α=0		
30	-.43(.013)	-.45(.036)	.50(.013)	.56(.036)
20	-.49(.017)	-.46(.037)	.49(.017)	.46(.037)
10	-.50(.022)	-.39(.038)	.48(.023)	.39(.040)
5	-.37(.030)	-.18(.035)	.54(.030)	.26(.034)
		N=100, α=0		
30	-.37(.043)	-.48(.118)	.64(.043)	.67(.114)
20	-.49(.047)	-.56(.096)	.49(.054)	.46(.112)
10	-.57(.065)	-.25(.142)	.49(.072)	.39(.146)
5	-.43(.092)	-.40(.104)	.63(.099)	.33(.113)

estimation. The subject parameters were estimated by maximizing the likelihood of the item responses given the CML estimates of α (Fischer, 1974, Equation 14.3.1, p. 251). Individuals with a zero or perfect score were excluded. The regression parameters β were estimated according to the model specified in (4), $\hat{\beta}$, and with the model specified in (11), $\tilde{\beta}$. Also the standard errors of $\hat{\beta}$ and $\tilde{\beta}$ were estimated. (A special Fortran-77 computer program was written to estimate $\hat{\beta}$ and its standard error, and can be obtained from the author.) The regression parameter $\tilde{\beta}$ and its standard error were obtained with SPSSX-module regression. The estimated subject parameter, $\hat{\theta}$, was the dependent variable, and x_1 and x_2 were entered simultaneously as independent predictor variables. The results are given in Table 1.

For the simulations with 20 or 30 items, the estimated regression parameters using the models in (4) and (11) were almost equivalent. Neither one was best in terms of being closest to the true values. In the simulations with ten or five items, $\hat{\beta}$ was closer to the true values of β than $\tilde{\beta}$. In all simulations the standard errors of $\hat{\beta}$ were smaller than the standard errors of $\tilde{\beta}$. The number of individuals, and the range of the item difficulties appeared not to affect the accuracy of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$. Simulations were also carried out with varying degree of correlation between the two predictors. The results of those simulations were not much different. It appeared that the regression parameters were estimated with equal accuracy as in the case of predictors being uncorrelated, but with larger standard errors.

Empirical Data

As an illustration of the logistic regression model we used item response data gathered in a study of the treatment of acute bronchitis with antibiotics. Only the necessary details will be discussed here. For an extended discussion of these data, see Zwinderman (in press), and Zwinderman, Verhey, Hermans, Kaptein, and Mulder (in press).

A sample of 434 general practitioners responded to a questionnaire of 11 simulated patients with varying number of clinical symptoms of acute bronchitis. All patients had coughing complaints, had fever or not, had audible rhonchi or not, and coughed up purulent sputum or not. Each practitioner was asked whether he or she would prescribe an antibiotic to each of the patients.

There is lack of agreement among practitioners with respect to the definition and the treatment of acute bronchitis. To obtain insight into this lack of agreement, we modeled the prescription probabilities according to the model specified in (4). The patients varied in seriousness as a function of the number and the kind of symptoms. The parameter α represented this variation. The practitioners varied with respect to their inclination to use antibiotics. The inclination to use antibiotics was a latent variable denoted by θ. The relation between the inclination to use antibiotics and several variables was studied with the model as specified in (4).

As predictor variables, we used (1) age, (2) sex (0 = male, 1 = female), (3) experience as a general practitioner, (4) the kind of practice in which the practitioner works, (5) the university where the general practitioner was trained, and (6) whether or not the practitioner had a pharmacy. There were 391 male practitioners, and 43 female. The mean age (standard deviation) of the practitioners was 43.4 (9.1). The average number of years experience was 14.9 (9.2). There were 223 practitioners working alone, 147 were associated with another specialty, and 64 worked in a group practice. This variable was recoded into two binary variables. There were practitioners of 8 Dutch universities, abbreviated as follows: UVA (61), VU (34), RUG (84), RUL (66), RUM (6), KUN (50), EUR (53), and RUU (80). This variable was recoded into seven binary variables. There were 83 practitioners who had a pharmacy.

The seriousness parameters of the patients were estimated according to the Rasch model with conditional maximum likelihood estimation. The CML estimates of $\boldsymbol{\alpha}$ were used as constants in the remainder. The standard deviation of θ was estimated as 1.68 using the approach of Andersen and Madsen (1977). The fit of the Rasch model was tested with Glas' CML-R_1 test statistic (Glas, 1988, 1989), Molenaar's U_i item statistic (Molenaar, 1983), and Andersen's conditional likelihood ratio statistic (Andersen, 1973) for the group of practitioners older than 40 and younger than 40. R_1 was 105 with 90 degrees of freedom ($p = 0.13$), the likelihood ratio statistic was 29 with 20 degrees of freedom ($p = 0.47$). For two patients, $|U_i|$ was larger than 2 (2.14 and 2.06). These test data showed that the Rasch model fit satisfactory.

The relation between the inclination to prescribe antibiotics and the predictor variables was estimated according to the model as specified in (4). The estimated

TABLE 2

Estimates of the Parameters of the Logistic Regression Model on the Inclination to Prescribe Antibiotics

Predictors	ß	se(ß)	z
Intercept	-0.35	.204	-1.72
Sex	-0.09	.044	-2.05
Age	-0.20	.154	-1.30
Experience	0.11	.153	0.72
Practice			
Alone	0.00	-	
Associated	-0.05	.045	-1.11
Group	0.02	.046	0.43
Pharmacy	0.18	.044	4.09
Training			
UvA	0.00	-	
VU	-0.09	.050	1.80
RUG	0.14	.059	2.37
RUL	0.09	.056	1.61
RUM	-0.17	.044	-3.86
KUN	-0.03	.054	-0.56
EUR	0.08	.054	1.48
RUU	0.11	.057	1.93
σ_ϵ	1.65	.19	

regression parameters are given in Table 2. For each parameter, the ratio (z) between the estimate and its standard error was calculated to test whether the parameter could be considered zero.

It appeared that female practitioners were less inclined to prescribe antibiotics than male practitioners, and those who had a pharmacy were more inclined to prescribe antibiotics. Several differences between practitioners trained at different universities were found. The multiple correlation between θ and $\boldsymbol{\beta}'\mathbf{x}$ was estimated as $(1 - \sigma_\varepsilon^2/\sigma_\theta^2)^{1/2} = 0.19$.

Conclusion

In general, ability parameters of latent trait models are not estimated consistently, and should not be used as a dependent variable in regression models. Direct estimation of the regression parameters with the logistic model yields better results in terms of accuracy and efficiency at least with small number of items. The logistic regression model is flexible and few assumptions concerning the predictors are needed. It can be readily applied in conjunction with any logistic latent trait model.

The structure of the Rasch model permits a set of items to be calibrated independently of the individuals in the calibration sample. This means that the item difficulties can be estimated independently of the regression model with conditional maximum likelihood estimation (CML). When the CML estimates of the item difficulties are used as known constants in the regression model, the computational effort is greatly reduced. Furthermore, most aspects of the regression model can thus be tested in the framework of the Rasch model. For any other latent trait model, the regression model cannot be estimated or tested independently of the calibration of the set of items. The measurement scale depends upon the regression model, and conversely.

In this paper it was assumed that all items were dichotomous and measured the same latent trait. The model as specified in (4) can be generalized to polytomous items in the same way as latent trait models can be generalized to polytomous items. For ordered categorical responses, the generalization remains a unidimensional model. For nominal responses, the generalization is a model with two or more dependent variables. The logistic regression model can also be generalized to sets of items that measure two or more different latent traits, also leading to a model with two or more dependent variables (see Zwinderman, in press).

Appendix

This appendix includes a derivation of the information matrix of the maximum likelihood estimators ($\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}$) of the parameters of the logistic regression model as specified in (4).

If $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}$ are the maximum likelihood estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ, the inverse of the information matrix evaluated at $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}$ is a consistent estimate of the covariance matrix of the parameter estimates. The information matrix is defined as the negative matrix with the second order partial derivatives of the logarithm of the conditional marginal likelihood as specified in (5). The entries of the information matrix are given by the following seven formulas. For a concise notation, we denote the posterior density function of ε given $(\mathbf{x}_v, \mathbf{y}_v)$, $p_v(\varepsilon|\mathbf{x}_v, \mathbf{y}_v)$, by p_v. The integrals run from minus infinity to infinity.

$$\frac{d^2 \log L_m}{d\beta_j \, d\beta_l} = \sum_{v=1}^{N} x_{vj}x_{vl}$$

$$\times \frac{\displaystyle\int_{-\infty}^{\infty} p_v \left(\sum_{i=1}^{k} (h_{vi})^2 - \sum_{i=1}^{k} h_{vi}(1-h_{vi}) \right) d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon - \left[\int_{-\infty}^{\infty} p_v \sum_{i=1}^{k} h_{vi} \, d\varepsilon \right]^2}{\left[\displaystyle\int_{-\infty}^{\infty} p_v \, d\varepsilon \right]^2}$$

$$\frac{d^2 \log L_m}{d\beta_j \, d\sigma^2} = \sum_{v=1}^{N} x_{vj}$$

$$\times \frac{-\int_{-\infty}^{\infty} p_v \sum_{i=1}^{k} h_{vi} \frac{1}{2} \frac{\varepsilon^2}{\sigma^4} \, d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon + \int_{-\infty}^{\infty} \sum_{i=1}^{k} h_{vi} \, d\varepsilon \int_{-\infty}^{\infty} p_v \frac{1}{2} \frac{\varepsilon^2}{\sigma^4} \, d\varepsilon}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2},$$

$$\frac{d^2 \log L_m}{d^2\sigma^2} = \frac{1}{2} N\sigma^{-4}$$

$$+ \sum_{v=1}^{N} \frac{\int_{-\infty}^{\infty} p_v \frac{\varepsilon^2}{\sigma^6} \left(\frac{1}{4} \frac{\varepsilon^2}{\sigma^2} - 1\right) d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon - \left[\int_{-\infty}^{\infty} p_v \frac{1}{2} \frac{\varepsilon^2}{\sigma^4} \, d\varepsilon\right]^2}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2},$$

$$\frac{d^2 \log L_m}{d^2\alpha_i} = \sum_{v=1}^{N} \frac{\int_{-\infty}^{\infty} p_v(h_{vi}^2 - h_{vi}(1 - h_{vi})) \, d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon - \left[\int_{-\infty}^{\infty} p_v h_{vi} \, d\varepsilon\right]^2}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2},$$

$$\frac{d^2 \log L_m}{d\alpha_i \, d\alpha_j} = \sum_{v=1}^{N} \frac{\int_{-\infty}^{\infty} p_v h_{vi} h_{vj} \, d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon - \int_{-\infty}^{\infty} p_v h_{vi} \, d\varepsilon \int_{-\infty}^{\infty} p_v h_{vj} \, d\varepsilon}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2},$$

$$\frac{d^2 \log L_m}{d\alpha_i \, d\beta_j} = \sum_{v=1}^{N}$$

$$\times \frac{\int_{-\infty}^{\infty} x_{vj} p_v \left(h_{vi}(1 - h_{vi}) - h_{vi} \sum_{i=1}^{k} h_{vi}\right) d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2},$$

$$+ \frac{\int_{-\infty}^{\infty} p_v h_{vi} \, d\varepsilon \int_{-\infty}^{\infty} x_{vj} p_v \sum_{i=1}^{k} h_{vi} \, d\varepsilon}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2}$$

$$\frac{d^2 \log L_m}{d\alpha_i \, d\sigma^2} = \sum_{v=1}^{N} \frac{\int_{-\infty}^{\infty} p_v h_{vi} \frac{1}{2}\frac{\varepsilon^2}{\sigma^4} d\varepsilon \int_{-\infty}^{\infty} p_v \, d\varepsilon - \int_{-\infty}^{\infty} p_v h_{vi} \, d\varepsilon \int_{-\infty}^{\infty} p_v \frac{1}{2}\frac{\varepsilon^2}{\sigma^4} d\varepsilon}{\left[\int_{-\infty}^{\infty} p_v \, d\varepsilon\right]^2}.$$

References

Abramowitz, M., & Stegun, I. A. (Eds.). (1964). *Handbook of mathematical functions*. New York: Dover Publications.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimates. *Journal of the Royal Statistical Society, Series B, 32*, 283–301.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123–140.

Andersen, E. B., & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 35*, 357–374.

Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika, 59*, 19–35.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459.

Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research, Vol. 1. The analysis of case-control studies*. Lyon: IARC Scientific Publications.

Dixon, W. J. (1985). *BMDP statistical software*. Berkeley: University of California Press.

Fischer, G. H. (1974). *Einführung in die Theorie Psychologischer Tests* [Introduction into the theory of psychological tests]. Bern: Verlag Hans Huber.

Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika, 53*, 525–546.

Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models*. Unpublished doctoral dissertation, Cito, Arnhem.

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology, 33*, 234–260.

Lord, F. M. (1980). *Applications of item response theory to practical test problems*. Hillsdale, NJ: Erlbaum.

Lord, F. M. (1984). *Maximum likelihood and bayesian parameter estimation in IRT* (RR-84-30-ONR). Princeton, NJ: Educational Testing Service.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

Molenaar, I. W. (1983). Some improved diagnostics for the Rasch model. *Psychometrika, 49*, 49–72.

Prentice, R. L., & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika, 66*, 403–411.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.

Rigdon, S. E., & Tsutakawa, R. K. (1983). Parameter estimation in latent trait models. *Psychometrika, 48*, 567–574.

van Houwelingen, J. C., & le Cessie, S. (1988). Logistic regression: A review. *Statistica Neerlandica, 42*, 215–232.

Zwinderman, A. H. (in press). A two stage Rasch model approach to dependent item responses: An application of constrained item response models. *Methodika*.

Zwinderman, A. H., Verhey, Th. J. M. Hermans, J., Kaptein, A. A., & Mulder, J. D. (in press). Application of the Rasch model for the prescription behavior of general practitioners. *Journal of Clinical Epidemiology*.

Manuscript received 1/31/89
Final version received 10/12/90