

# ANALYSING COGNITIVE TEST DATA: DISTRIBUTIONS AND NON-PARAMETRIC RANDOM EFFECTS

GRACIELA MUNIZ TERRERA, ARDO VAN DEN HOUT, R.A. RIGBY,  
AND D.M. STASINOPOULOS

**ABSTRACT.** An important assumption in many linear mixed models is that the conditional distribution of the response variable is normal. This assumption is violated when the models are fitted to an outcome variable that counts the number of correctly answered questions in a questionnaire. Examples include investigations of cognitive decline where models are fitted to Mini-Mental State Examination (MMSE) scores, the most widely used test to measure global cognition. MMSE scores take integer values in the 0-30 range, and its distribution has strong ceiling and floor effects. This paper explores alternative distributions for the outcome variable in mixed models fitted to MMSE scores from a longitudinal study of ageing. Model fit improved when a beta-binomial distribution was chosen as the distribution for the response variable.

## 1. INTRODUCTION

The Mini Mental State Examination (MMSE, [1]) is a test that was originally designed to provide a brief and standardised assessment of mental status in psychiatric patients. However, it has become the most widely used test to assess global cognition in older individuals. It consists of a set of questions devised to assess cognitive domains such as memory, language, attention and orientation for time and place. Each of the individual questions is scored as 1 if correctly answered and 0 if not, and the final MMSE score, an integer in the range 0-30, is calculated adding up these scores. High values indicate good cognition. Individuals with scores in the 25-30 range are said to have normal cognitive function, individuals with scores

---

*Key words and phrases.* cognitive test; beta binomial; random effects models.

between 24 and 10 are said to be cognitive impaired, whilst scores below 10 are indicative for dementia.

The MMSE is known to have a skewed distribution and strong ceiling and floor effects [2]. Ceiling effects happen when tests are relatively easy, and a substantive group of individuals obtain maximum or near maximum scores and cannot demonstrate their true ability [3]. Floor effects are similar to ceiling effects but at the other end of the cognitive spectrum. In the presence of floor and ceiling effects, variability amongst individuals is underestimated as the identification of differences amongst high or poor scorers is limited.

Examinations of cognitive decline are often conducted by fitting linear mixed models [4]. Linear mixed models take into account correlated data, are flexible in handling missing data and permit the investigation of between and within individual changes. An assumption often made when fitting these models is that, conditional on the random effects, the response variable is normally distributed. When linear mixed models are fitted to MMSE scores, this assumption is likely to be violated due to the discrete nature of the sum scores, their skewed distribution and floor and ceiling effects. Proposals found in the literature to satisfy the normality assumption of linear mixed models fitted to MMSE scores include transformations of the scores such as  $\sqrt{31 - MMSE}$  [5] and  $\log(31 - MMSE)$  [6]. Others have suggested the analysis of a composite measure obtained from the combination of z-scores calculated using results from more than one cognitive test [7]. However, difficulties arising from the interpretation of results from models fitted to transformed scores and the fact that in many studies researchers lack results from more than one cognitive test, make these proposals unappealing to many investigators. More importantly, these solutions do not acknowledge the discrete nature of the cognitive scores.

Linear regression and other regression-based models such as generalised linear models [8] and generalised additive models [9] assume an exponential family distribution for the response variable. The mean  $\mu$  of the response variable is modelled as

a function of explanatory variables, and the variance depends on a constant dispersion parameter or on  $\mu$ . Extensions to mixed-effects models such as linear mixed models, generalised linear mixed models and generalised additive mixed models maintain the exponential family distribution assumption for the response variable and only seldom allow the modelling of the variance as a function of explanatory variables.

Generalised additive models for location, scale and shape (GAMLSS, [11]) were developed to incorporate a wide range of regression-based models, and, in addition, to relax the assumption of the exponential family distribution for the response variable. GAMLSS allow regression equations for all the parameters of the distribution for the response variable. The equations may include linear functions of the explanatory variables, additive functions, and random effects. GAMLSS models have been implemented in the freely available R package `gamlss` [11], see also <http://gamlss.org>.

In this paper we propose to investigate the choice of the conditional distribution of the response variable in mixed models fitted to longitudinal MMSE scores. We will show that replacing the often used continuous normal distribution by the discrete binomial and beta-binomial distributions leads to improved statistical inference. We envisage that similar improvement can be attained in many other applications in medical statistics where the response variable in a mixed regression model is discrete and its conditional distribution is poorly approximated by a normal distribution.

The paper is organised as follows. Section 2 introduces the data from the Cambridge City over 75 Cohort Study. In Section 3, models, estimation, and model fit are discussed. Section 4 presents the data analysis. Section 5 is the discussion.

## 2. DATA

The Cambridge City over 75 Cohort Study (CC75C) is a prospective study of a representative population sample of around 2600 people aged 75 years and older

at the beginning of the study, in 1985. The study was designed with the aim of conducting research on topics such as prevalence and incidence of dementia, and the investigation of risk factors for cognitive decline, neuropsychology and depression. Originally, all patients registered in six general practices in the city of Cambridge who were aged 75 years old and over on April 1st, 1985, were considered. One in three patients from another general practice were also sampled. In this practice, with the purpose of avoiding sampling bias, it was specially verified that the age of those sampled and those not sampled were very similar. Patients living in residential and nursing homes were not excluded. Selected individuals were then invited to take part in the study by their doctors. The 2611 individuals who accepted the invitation constituted 95% of those invited. All remaining participants had a screening interview carried out by a trained interviewer where patients' details, family contacts, health status and use of health services were recorded.

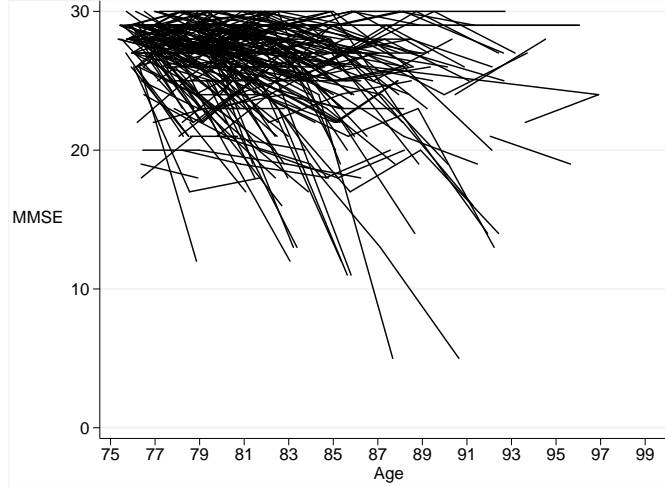
The screening interview was followed by a more detailed clinical interview of all individuals scoring 23 or less on the MMSE and a third of those with scores of 24 and 25 points. Further waves of interviews were carried out to establish incidence of dementia. After the baseline interview, survivors have had up to six follow-up interviews conducted, on average, 2, 7, 9, 11, 13 and 17 years later. Marital status, social networks, use of social services, physical abilities, and other sociodemographic information was collected (see the study's webpage [www.cc75c.cam.ac.uk](http://www.cc75c.cam.ac.uk) for more details) at baseline and other survey interviews. Of the 2075 study participants with valid scores at baseline, 65% were women, and 50% had a physical disability at baseline.

For our investigation, we examined data from a randomly selected sample of 360 men aged between 75 and 80 years old at baseline. Sixty-five per cent of the men left school aged younger than 15 years old. At the different interviews, the number of men in the sample with MMSE scores of 30 were 24, 16, 9, 7, 3 and 1 individuals, respectively. Scores above 27 MMSE points were obtained by 248, 127, 60, 39, 12, 6 and 1 person, respectively, at the different study interviews. Table 1 shows further

TABLE 1. Number of men interviewed, mean (and standard deviation) of MMSE scores and age of men at each of the seven study interviews

Years after initial interview	N	MMSE	Age
Baseline	360	26.7 (2.9)	77.5 (1.3)
2	213	26.2 (2.9)	79.8 (1.3)
7	133	24.9 (4.3)	83.4 (1.3)
9	80	25.0 (4.4)	86.3 (1.3)
11	38	23.5 (5.7)	89.7 (1.4)
13	13	22.7 (8.0)	93.0 (1.0)
17	2	26.5 (3.5)	96.4 (0.6)

FIGURE 1. MMSE trajectories plotted as a function of age in the sample of CC75C male study participants aged 75-80 years old at baseline



characteristics of the MMSE scores and age of the men in the sample examined over the study interviews.

Figure 1 shows MMSE scores plotted as a function of age for all men in the sample and illustrates the features of the MMSE distribution as it shows the substantial number of individuals with high MMSE scores.

### 3. METHODS

3.1. **Models.** GAMLSS are formulated with four parameters for the conditional distribution of the response. As an example, the standard linear regression model

is included and has two distribution parameters, one for the mean, and one for the variance. Generalised linear models are also included. The binary logistic model, for example, has a logit link and one parameter for the binomial distribution for the binary response.

We start by formulating GAMLSS in a general way. Let  $y_i$ ,  $i = 1, \dots, n$ , be conditional independent observations with density function  $f(y_i|\boldsymbol{\theta}^i)$ , with  $\boldsymbol{\theta}^i = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4})$  a vector of four distribution parameters (location, scale, and two shape parameters). Let the index  $k$  correspond with these four distribution parameters. For each  $k = 1, \dots, 4$  a monotonic link function  $g_k(\cdot)$  can be specified relating the distribution parameter vector  $\boldsymbol{\theta}_k = (\theta_{1k}, \dots, \theta_{nk})$  to predictor  $\boldsymbol{\eta}_k = (\eta_{1k}, \dots, \eta_{nk})$ . The four regression equations in GAMLSS are given by

$$(1) \quad g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{l=1}^{L_k} \mathbf{Z}_{lk} \boldsymbol{\gamma}_{lk} \quad \text{for} \quad k = 1, 2, 3, 4,$$

where  $\boldsymbol{\beta}_k$  is a vector with regression coefficients for the known design matrix  $\mathbf{X}_k$ ,  $\mathbf{Z}_{lk}$  is a known  $n \times q_{lk}$  design matrix and  $\boldsymbol{\gamma}_{lk}$  is a  $q_{lk}$  dimensional random variable the distribution of which is parameterised by a  $q_{lk} \times q_{lk}$  symmetric matrix that may depend on hyperparameters.

The vectors  $\boldsymbol{\gamma}_{jk}$  could be combined in a single vector with a single matrix, which would remove the summation in (1). However, in line with the original model formulation in [11], we maintain the formulation in (1), which also better shows that different types of random-effects terms can be included in the model.

To obtain the standard linear regression model,  $\boldsymbol{\theta}^i = (\theta_{i1}, \theta_{i2}) = (\mu_i, \sigma_i)$ ,  $g_1(\cdot)$  is the identity link,  $g_2(\cdot)$  is the log link,  $\boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1$ , and  $\boldsymbol{\eta}_2 = \beta_2$ , and  $f(y_i|\boldsymbol{\theta}^i)$  is the density of a normal distribution. To obtain the binary logistic model,  $\boldsymbol{\theta}^i = \theta_{i1} = \mu_i$ ,  $g_1(\cdot)$  is the logit link,  $\boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1$ , and  $f(y_i|\boldsymbol{\theta}^i)$  is the probability mass function of a binomial distribution.

A semiparametric additive model can be expressed by choosing  $\mathbf{Z}_{jk} = \mathbf{I}_n$ , where  $\mathbf{I}_n$  is a  $n \times n$  identity matrix, and  $\boldsymbol{\gamma}_{jk} = h_{jk}(\mathbf{x}_{jk})$ , where  $\mathbf{x}_{jk}$  for  $j = 1, 2, \dots, J_k$  are

vectors of length  $n$ . The function  $h_{jk}(\mathbf{x}_{jk})$  evaluates the unknown function  $h_{jk}$  at the design vector  $\mathbf{x}_{jk}$ . For more details and additional examples, see [11].

In this paper, we consider random-effects GAMLSS where the regression equation for the location  $\boldsymbol{\theta}_1 = \boldsymbol{\mu}$  is given by  $g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{l=1}^2 \mathbf{Z}_{l1}\boldsymbol{\gamma}_{l1}$ . Four different distributions for the response variable will be investigated: the normal distribution (Model 1), the *Student t*-family distribution (Model 2), and two discrete distributions, namely the binomial distribution (Model 3) and the beta-binomial distribution (Model 4).

In Model 1,  $g_1(\cdot)$  is the identity link, and  $g_2(\cdot)$  is the log link for the scale parameter  $\boldsymbol{\theta}_2 = \sigma$ . In Model 2 the same links are used plus one extra log link for the second scale parameter. In Models 3 and 4,  $g_1(\cdot)$  is the logit link. In Model 4,  $g_2(\cdot)$  is the log link for the scale parameter of the beta-binomial distribution. Note that there is no scale parameter for the binomial distribution.

We elucidate the choice of Models 1 up to 4. Model 1 is the starting point as this model is often used in analyses of MMSE scores and we want to show that we can do better. Model 2 with the *t*-family distribution was chosen in consideration of the fact that the data may exhibit heavier tails than in the normal distribution. The normal and *t*-family distributions in Models 1 and 2 are likely to produce fitted values outside the range of the MMSE due to the ceiling effects. This introduces a dependence between residuals and fitted values which is a violation of the model assumptions. With respect to practice, note that predicted values of MMSE outside the original range 0-30 do not have a clear interpretation.

Models 3 and 4 appreciate the discrete nature of the response variable. The beta-binomial distribution can be used if there is overdispersion that is not captured by the binomial. Hence the extra scale parameter in the beta-binomial. This distribution is a mixture distribution. First assume that the response variable  $Y$  follows a binomial distribution,  $\text{Bin}(m, \pi)$ , with  $m = 30$ . Then, assume that  $\pi$  has a beta distribution with parameters  $\alpha, \beta > 0$ . Marginally, averaging over the beta

distribution for  $\pi$ ,  $Y$  follows a beta-binomial distribution with density function

$$p(y|\alpha, \beta) = \binom{m}{y} \frac{B(\alpha + y, m + \beta - y)}{B(\alpha, \beta)},$$

where  $B(\alpha, \beta)$  is the beta function. With definitions  $\mu = \alpha/(\alpha + \beta)$  and  $\sigma = 1/(\alpha + \beta)$ , the beta-binomial distribution for  $Y$  has first and second moments given by  $E(Y) = m\mu$  and  $var(Y) = m\mu(1 - \mu)[1 + (m - 1)\sigma/(1 + \sigma)]$ , respectively.

The standard choice of the distribution of random effects in a regression equation is a multivariate normal. Here, we relax this parameteric assumption and estimate the distribution by using a non-parametric mixing distribution. For generalised linear models, this approach is discussed in [14] and [16]. Note that Model 4 is not a generalised linear model since the beta-binomial model is not a member of the exponential family.

**3.2. Statistical inference.** Models were estimated by non-parametric maximum likelihood (NPML) using the EM algorithm [16] as implemented in the GAMLSS software. In NPML, the non-parametric mixing distribution for the random effects is discrete on an unknown finite number  $K$  of mass points  $\mathbf{z}_k$ , with masses  $p_k$ . The likelihood is given by

$$p(\mathbf{y}) = \prod_{i=1}^n \sum_{k=1}^K p_k p(y_i | \mathbf{z}_k)$$

Number of components  $K$ , mass points  $\mathbf{z}_k$ , and masses (or mixture proportions)  $p_k$  are unknown and are estimated from the data. Although  $K$  is unknown, it is treated as fixed when estimating the models. The EM algorithm consists of an M-step in which the  $\mathbf{z}_k$  and fixed-effects are estimated, and the E-step in which the probability that an observation comes from component  $k$  is estimated for all observations and all  $k$ . Details of the estimation can be found in [11] and [16].

The optimal  $K$  is chosen by comparison of global deviance (minus two times the maximised loglikelihood function) and the Akaike's Information Criterion (AIC) [15]. These comparisons are made heuristically as large-sample normality of the



likelihood function is an assumption that fails for mixture likelihoods, see [16], Section 7.6. Only large differences in the global deviances and the AIC will be taken into account.

Since there is no satisfactory way of statistically testing competing models, assessment of residuals is used as an additional tool for model comparison. Residuals are derived from fitted values, and the latter can be assessed at two levels in our NPML models. We distinguish marginally fitted values that are computed using the marginal mixture proportions  $p_1, \dots, p_K$ , and within-group fitted values that are computed using individual-specific mixture probabilities.

The term *within-group* is used in a similar way in parametric mixed-effects models (cf. [23], Section 4.3). In our case, the group is defined by the observations within one individual.

Individuals are indexed by  $j$ , and the number of observations for individual  $j$  by  $n_j$ . The mixture probability is denoted  $w_{jk}$  and it is the probability that the observations  $y_{j1}, \dots, y_{jn_j}$  for individual  $j$  come from component  $k$ . Define  $\hat{m}_{jk} = \prod_{r=1}^{n_j} p(y_{jr} | \hat{\mathbf{z}}_k)$ , where  $p(y_{jr} | \hat{\mathbf{z}}_k)$  is the density defined by the NPML model given the estimated mass points  $\mathbf{z}_k$ . The estimator of  $w_{jk}$  is given by

$$\hat{w}_{jk} = \frac{\hat{p}_k \hat{m}_{jk}}{\sum_{l=1}^K \hat{p}_l \hat{m}_{jl}},$$

see [16], Section 9.3. The corresponding within-group fitted values are  $\hat{y}_{jr} = \sum_{k=1}^{\hat{K}} \hat{w}_{jk} g_1^{-1}(\hat{\eta}_{jrk})$ , where  $\hat{\eta}_{jrk}$  is the linear predictor defined by component  $k$ , and  $g_1(\cdot)$  is the first link function in the GAMLSS formulation. The marginally fitted values are obtained by replacing  $\hat{w}_{jk}$  with  $\hat{p}_k$ .

The within-group fitted values and the corresponding residuals are not provided by the current GAMLSS implementation in the package `gamlss`, but the additional code can be obtained by contacting the authors.

For Models 3 and 4, randomised quantile residuals [24] are defined using the observed values and the within-group fitted values. Due to the link function in these models, assessing directly the difference between observed values and the

within-group fitted values is of limited value as there is no obvious distribution for these differences. The randomised quantile residuals, on the other hand, should follow a standard normal distribution if the model is correct.

#### 4. ANALYSIS

The location of the chosen distribution in GAMLSS for the MMSE scores was modelled using predictors with linear and quadratic functions of age. Intercept and the slope for the linear function of age are modelled as random effects using NPML.

For linear mixed models with parametric distributions of random effects, comparing linear and quadratic functions of age in a study of cognitive change was also undertaken in [12]. The analyses in this section extends [12] in two ways: different choices for the distribution of the response, and relaxing the parametric distribution for the random effects by replacing it by a discrete mixture distribution.

To be able to interpret the intercept in Models 1 and 2 as the predicted MMSE at age 75, i.e., the minimum age at the beginning of the study, we subtracted 75 from each individual's age at each interview. Following evidence of the effect of education on cognitive function in old age [13], we adjusted the model's intercept and slope (rate of change per extra year of older age) for education. With this aim, we created a binary variable that took the value of 1 if the individual had left school aged younger than 15 years old (average age at which individuals left school in this study) and 0 otherwise.

For brevity, we only present the formulation of the models where observed scores are linked to a linear function of age. The extension to quadratic function with a fixed effect for the quadratic term is straightforward.

Model 2 has an extra equation with the log link for the second scale parameter, namely  $g_3(\nu_j) = \log(\nu_j) = \alpha_\nu$ .

For Model 4, the equations for the location and scale are given by

$$\begin{aligned} g_1(\mu_{jt}) &= \text{logit}(\mu_{jt}) = (\beta_{1j} + \beta_2 \text{Educ}_j) + (\beta_{3j} + \beta_4 \text{Educ}_j) \text{Age}_{jt} \\ g_2(\sigma_j) &= \log(\sigma_j) = \beta_\sigma. \end{aligned}$$

where  $\beta_{1j}$  and  $\beta_{3j}$  are the random effects, and  $Age_{jt}$  as before. Model 3 has the same equation for the location, but has no scale parameter.

Although we have use the same notation for the parameters in the equations, these parameters should of course be interpreted within each of the models.

All models were fitted with the number of mass points  $K$  increasing from 4 up to a maximum of 10 . We limited the number of components fitted to 10 as a model with more than 10 mixture components would have resulted in a model with several components with very few individuals and low probabilities. Consistently, Akaike Information Criterion (AIC) and BIC values corresponding to models where the location was linked to a linear function of age were lower than those corresponding to models where the scores where modelled as a quadratic function of age. AIC is a measure of model fit calculated as  $AIC = -2Ln(lik) + 2K$ , where  $lik$  is the maximised likelihood and  $K$  the number of model parameters whilst the Bayesian Information Criterion (BIC) are calcaulted as  $BIC = -2Ln(lik) + Klog(N)$  with  $K$  as before and  $N$  the number of individuals. Hence, given the AIC and BIC values,we omit results from the quadratic models.

Table 2 shows global deviance ( $-2 \times \log\text{likelihood}$ ), number of parameters, AIC and BIC values and the number of iterations required to achieve convergence of the EM algorithm as implemented in `gamlss` for each of the models fitted. For example, Model 1 with the normal distribution and  $K = 4$  has 14 parameters: 4 mass points for the random intercept, 4 mass points for the random slope for age, 2 fixed effects for education, 3 independent parameters for the mixture proportions, and 1 parameter for the variance.

For the continuous distributions in Table 2 there is a considerable drop in global deviance when  $K$  is increased from 4 to 6. However, the real gain is by switching to a discrete distribution. According to the AIC, an eight-component beta-binomial model fitted the data best. Figure 2 shows the estimated marginal means of the  $K = 8$  classes for this model. Note that the curve that has a steep decline has a very small weight, namely 0.005. The BIC value corresponding to an eight-component

TABLE 2. Summary statistics for NPML Models 1 up to 4 with increasing number of mixture components.

Distribution	K	Number of iterations	Number of parameters	Global Deviance	AIC	BIC
Normal	4	62	14	4607.2	4235.2	4701.6
	6	175	20	4584.5	4624.5	4719.4
	8	28	26	4556.0	4608.0	4731.3
	10	86	32	3993.0	4057.0	4208.4
<i>t</i> -Family	4	10	15	3959.0	3989.0	4060.0
	6	16	21	3905.7	3947.7	4047.1
	8	23	27	3867.2	3927.2	4069.1
	10	13	33	3867.2	3939.3	4106.6
Binomial	4	12	13	3846.7	3872.7	3934.3
	6	34	19	3764.1	3802.1	3991.9
	8	21	78	3827.4	3877.4	3995.7
	10	23	31	3737.5	3799.5	3946.1
Beta	4	20	14	3816.0	3844.0	3910.2
Binomial	6	32	20	37740	3814.2	3908.8
	8	27	26	3741.5	3793.5	3916.5
	10	21	32	3734.1	3798.1	3949.5

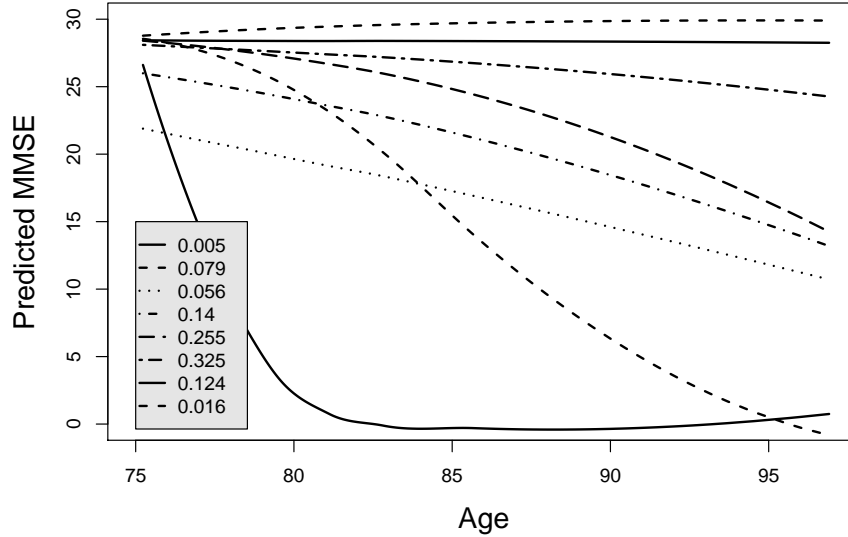
beta-binomial model is not the lowest, although the difference in BIC values between the eight component and the models with the next lowest BIC values is not large.

The advantage of using discrete distributions is supported by the inspection of the residuals plots in Figure 3. The residuals for the normal model with  $K = 10$  show a dependency on fitted values: for predicted values in the range 28-29, residuals are smaller than for instance in the range 20-25. It is also clear that the residuals are not standard normally distributed.

For the randomised quantile residuals for the binomial with  $K = 10$  and the beta-binomial model with  $K = 8$ , Figure 3 show that residuals are standard normally distributed and the plots illustrate good fit for these models.

Estimated masses, and estimated mass points for the intercept and slope are presented in Table 3 for the best fitting models for each of the four distributions considered. Specifically, Table 3 shows results from a 10 components normal and

FIGURE 2. Estimated marginal means for the beta-binomial model with  $K = 8$  components and the corresponding mixtures proportions. The two top lines correspond to the bottom two proportions. Means smoothed with respect to the effects of education.



binomial models and 8 components Student's  $t$  and beta-binomial models. The table shows the adjustment for the intercept and slope for each of the mixture components. For instance, to obtain the intercepts of the different components of the beta-binomial model, the values 1.40,  $-0.70$ , and other values listed under the column  $\beta_{1j}$  of the results corresponding to the beta-binomial model, should be added to 2.06. As a result, the intercept of the second component would be calculated as  $3.46 = 2.06 + 1.40$ , and the intercept of the third component would be  $1.36 = 2.06 - 0.70$ . A similar procedure should be used to calculate the slope of the different components of the models.

FIGURE 3. Residuals for the normal model with  $K = 10$ , and randomised quantile (NQ) residuals for the binomial with  $K = 10$  and the beta-binomial model with  $K = 8$ , respectively. Corresponding QQ-plots at the right-hand side.

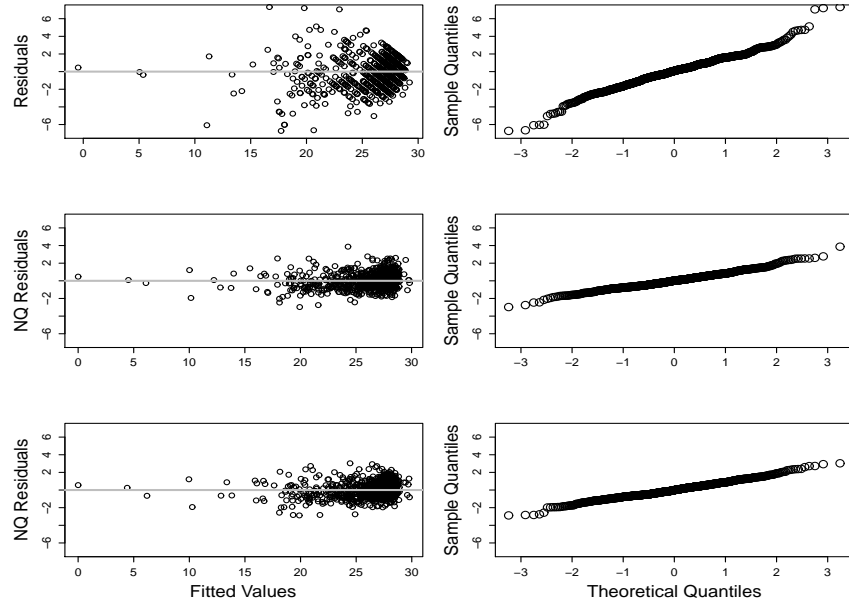


TABLE 3. Estimated NPML parameters (standard errors) and probabilities of best fitting models for each of the normal, Student's  $t$ , binomial and beta-binomial models

Normal( $K = 10$ )			t-family ( $K = 8$ )		
Masses ( $p_k$ )	Mass points ( $z_k$ )		Masses ( $p_k$ )	Mass points ( $z_k$ )	
	$\alpha_{1j}$	$\alpha_{3j}$		$\alpha_{1j}$	$\alpha_{3j}$
0.002	49.31(11.59)	-12.54(3.68)	0.028	20.15(0.48)	-0.13(0.07)
0.009	-25.29(11.63)	11.19(3.69)	0.065	9.39(0.56)	-1.35( 0.09)
0.069	-16.86(11.60)	10.85(3.69)	0.058	5.52(0.57)	-0.20(0.09)
0.040	-27.76(11.60)	12.21(3.69)	0.172	8.94(0.50)	-0.42(0.07)
0.089	-22.80(11.60)	12.03(3.68)	0.340	8.94(0.48)	-0.11(0.07)
0.167	-19.58(11.60)	11.93(3.68)	0.234	8.73(0.49)	0.06(0.07)
0.250	-20.38(11.59)	12.29(3.68)	0.099	9.17(0.53)	0.14(0.08)
0.365	-20.67(11.59)	12.49(3.68)	0.004	9.17(4.05)	0.14(0.50)
0.004	-20.67(11.71)	12.49(3.69)	-	-	-
0.000	-20.68(60.50)	12.49(8.79)	-	-	-
$\alpha_2$	-0.90(0.24)		$\alpha_2$	-0.78(0.18)	
$\alpha_4$	0.013(0.03)		$\alpha_3$	-0.01(0.02)	

Binomial ( $K = 10$ )			Beta-Binomial ( $K = 8$ )		
Masses ( $p_k$ )	Mass points ( $z_k$ )		Masses ( $p_k$ )	Mass points( $z_k$ )	
	$\beta_{1j}$	$\beta_{3j}$		$\beta_{1j}$	$\beta_{3j}$
0.004	2.06( 0.46)	-0.86(0.20)	0.005	2.06(0.47)	-0.85(0.21)
0.003	1.38(0.86)	-0.02(0.31)	0.079	1.40(0.48)	0.53(0.21)
0.048	-0.76(0.46)	0.78(0.20)	0.056	-0.70(0.47)	0.76(0.21)
0.086	1.33(0.47)	0.54(0.20)	0.140	0.19(0.47)	0.74(0.21)
0.129	0.07(0.46)	0.75(0.20)	0.255	1.22(0.47)	0.70(0.21)
0.245	1.20(0.46)	0.70(0.20)	.325	1.01(0.47)	0.78(0.21)
0.354	1.02(0.46)	0.78(0.20)	0.124	1.20(0.49)	0.83(0.21)
0.113	1.25( 0.48)	0.84(0.20)	0.016	1.42(0.84)	0.98(0.24)
0.014	1.45(0.82)	0.99(0.24)	-	-	-
0.000	1.54(4.03)	0.99(0.75)	-	-	-
$\beta_2$	-0.52(0.07)		$\beta_2$	-0.54(0.08)	
$\beta_4$	0.02(0.01)		$\beta_4$	0.03(0.01)	

For predictive purposes, it may be useful to describe model results obtained using an alternative representation. For instance, for a man aged 75 at study entry, with good education, results obtained from the normal model may be expressed as:

$$\text{logit}(\mu_{jt}) = \left\{ \begin{array}{ll} 2.06 - 0.85\text{Age}_{jt} & p = 0.005, \\ 3.46 - 0.32\text{Age}_{jt} & p = 0.079, \\ 1.36 - 0.09\text{Age}_{jt} & p = 0.056, \\ 2.25 - 0.11\text{Age}_{jt} & p = 0.140, \\ 3.28 - 0.15\text{Age}_{jt} & p = 0.255, \\ 3.07 - 0.15\text{Age}_{jt} & p = 0.325, \\ 3.26 - 0.02\text{Age}_{jt} & p = 0.124, \\ 3.48 + 0.13\text{Age}_{jt} & p = 0.016, \end{array} \right.$$

Further results from the model indicate that individuals with poor education have lower cognitive function than individuals with better education ( $\hat{\beta}_2 = -0.54$ , with estimated standard error = 0.08, p-value <0.00) and faster rate of decline ( $\hat{\beta}_4 = 0.03$ , with estimated standard error 0.01, p-value=0.002). The logarithm of the variance was estimated at  $-6.02$  (0.69).

With regard to interpretation it is important to realise that the binomial distribution for the MMSE is an approximation. It is not the case that the MMSE questions are independent Bernoulli trials. However, we can use a fitted binomial model to get a general idea of the effect of age by assuming that the questions are independent, and that  $\mu$  is the probability of a correct answer. This approximation is the same for the beta-binomial model and will be used in what follows.

Of the eight mixture components identified in the beta-binomial model, two (first and eighth) have very low probability. The first component shows that the estimated odds of answering a question correctly multiplies by 0.42 ( $= \exp(-0.85)$ ) per extra year of age after age 75 years with probability 0.005. The eighth component shows that the odds of correctly answering a question increases by 14 per cent per extra year of older age with very low probability (0.016). These two components



describe trajectories of individuals whose MMSE either dropped markedly (first component) or show some improvement (eighth component).

Components four, five and six suggest that, with total probability of 0.72 ( $= 0.14 + 0.26 + 0.32$ ), there is about 12% decrease in the odds of correctly answering a question per extra year of older age whilst the second mixture component indicates that these odds decrease by 28% with probability 0.07. The remaining components capture MMSE scores of individuals for whom the odds of correctly answering a question do not decrease very markedly with increasing age.

## 5. DISCUSSION

With the aim of analysing longitudinal data for the Mini-Mental State Examination (MMSE), we fitted generalised additive models for location, scale and shape (GAMLSS). In these models, a non-parametric distribution was used for the random effects. We considered four distributions for the MMSE as response variable: normal,  $t$ -family, binomial and beta-binomial. Our results show that the beta-binomial distribution yielded the best fit to the data.

Traditional approaches to examine change in cognitive function consist of fitting linear random-effects models using the normal distribution for the response. However, due to the distributional properties of the MMSE such as floor and ceiling effects and its skewed distribution, these models may not be appropriate. Researchers could initially be inclined to consider discrete mixture models to account for non-normal random effects, as conditional upon components, a normal distribution of the random effects may be defensible. However, within components ceiling effects might still operate and the bounded sample space for the outcome will still be present.

Tobit growth models have been proposed to account for ceiling and floor effects [17, 18]. Tobit models are extensions of growth models for censored outcome variables. Wang [19] compared the performance of tobit growth models to listwise deletion and to a method where ceiling data were considered as missing data to

deal with ceiling effects and concluded that tobit growth models performed best in terms of making best use of ceiling data information. An undesirable effect of tobit models is that they may produce values outside the range of the response variable. For instance, in the case of the MMSE, tobit growth models may fit or predict a score larger than 30 and such a score would lack meaningful interpretation in terms of cognitive function.

An alternative to using the MMSE sum score as the response variable is using the scores on the individual MMSE questions and fit item response theory (IRT) models [25], where the performance on the test is explained by a latent ability which is assumed to follow a normal distribution across individuals. The advantage is that the data are used at a more informative level, and that the ceiling effects in the MMSE are not a problem with regard to the choice of distributions, see Fox [26], Chapter 6, for an application. The disadvantage is that the ability which underlies the MMSE performance is latent (and on an arbitrary scale) and that straightforward interpretation of estimated ability and/or covariate effects is not possible.

The choice of a discrete distribution for the MMSE scores means that fitted and predicted scores are in the range of the original scale. Binomial models fitted the data better than models with the normal distribution or a  $t$ -family distribution. Further improvement in fit was obtained using the beta-binomial model which has an extra scale parameter to capture heterogeneity that cannot be described by the binomial distribution.

Non-parametric maximum likelihood (NPML) estimation for mixed models was initially proposed for generalised linear models in the presence of overdispersion. The estimation of overdispersed models with normal random effects requires Gaussian quadrature because the likelihood function does not have a closed form. However, results are sensitive to the parametric form of the mixing distribution of the random effects [20]. To overcome this limitation, NPML uses a discrete distribution on a finite number of mass points [4, 22].

Aitkin [14] discussed advantages and limitations of NPML. Advantages include the fact that it is relatively easy to increase the number of mass points, and that the mass points do not need to be restricted to a grid. Limitations include the fact that, in general, the mixture components do not have substantive meaning but are a discrete version of the variation usually regarded as continuous.

An advantage of the work in the current paper is that model fitting can be undertaken using a freely available and very flexible software package. The GAMLSS package allowed us to examine a series of models that are not within the standard specification of the most commonly used statistical packages.

Possible extensions of our work include investigation of more complex models for scale, skew and kurtosis parameters.

#### REFERENCES

- [1] Folstein M F, Folstein SF, McHugh PR. Mini-Mental State: a practical method for grading the state of patients for the clinician. *Journal of Psychiatric Research* 1975; **12**: 189-198.
- [2] Folstein M F, Folstein SF, McHugh PR. Key papers in geriatric psychiatry. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *International Journal of Geriatric Psychiatry* 1998; **13(5)**:285-294.
- [3] Uttl,B. Measurment of individual differences. lessons from memory assessment in research and clinical practice. *Psychological Science* 2005; **16(6)**:460-467.
- [4] Laird N, Ware J. Random effects models for longitudinal data. *Biometrics* 1982; **38**: 963-974.
- [5] Proust C, Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random effects. *Computational Methods and Programs in Biomedicine* 2005; **78(2)**:165-173.
- [6] Huppert F.A.,Cabelli S.,Matthews F.E. and MRC CFAS.Brief cognitive assessment in a UK population sample distributional properties and the relationship between the MMSE and an extended mental state examination. *BMC Geriatrics* 2005;**5**:7
- [7] Wilson RS, Beck TL, Bienias JL, Bennet DA. Terminal cognitive decline: accelerated loss of cognition in the last years of life. *Psychosomatic Medicine* 2007; **69(2)**:131-137.
- [8] Nelder JA, Wedderburn RWM. Generalized linear models. *J.R. Statistic. Soc. A*;**135**: 370-384.
- [9] Hastie TJ, Tibshirani R.J. Generalized additive models. 1990. Chapman and Hall. London UK.
- [10] Nelder J, Wedderman R. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A* 1972; **135**:370-384.

- [11] Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Applied Statistics* 2005; **54**:507-554.
- [12] Muniz-Terrera G, Matthews FE, Brayne C. A comparison of parametric models for the investigation of the shape of cognitive change in the older population. *BMC Neurology* 2008.
- [13] Muniz-Terrera G, Matthews F, Denning T, Huppert FA, Brayne C; CC75C Group. Education and trajectories of cognitive decline over 9 years in very old people: methods and risk analysis. *Age Ageing* 2009;**38**(3):277-82.
- [14] Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 1999;**55**:117-128.
- [15] Akaike, H. Factor analysis and AIC. *Psychometrika* 1987;**52**:317-332.
- [16] Aitkin, M., Francis, B, Hinde, J and Darnell, R. Statistical Modelling in R. Oxford statistical Science Series. 2009.
- [17] Tobin J. Estimation of relationships for limited dependent variables *Econometria* 1958; **26**(1):24-36.
- [18] Twisk J., Rijmen F. Longitudinal tobit regression: a new approach to analyze outcome variables with floor and ceiling effects. *Journal of Clinical Epidemiology* 2009; **62**:953-958.
- [19] Wang L, Zhang Z, McArdle J, Salthouse T. Investigating ceiling effects in longitudinal data analysis. *Multivariate behavioral Research* 2010; *43*(93): 476-496.
- [20] Heckman J, Wood AT. A method for minimizing the impact of distributional assumptions in econometric models of duration. *Econometrica* 1987; *52* 271-320.
- [21] Laird N. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* 1978; *73*,805-811.
- [22] Lindsay BG. The geometry of mixture likelihoods, part I: a general theory. *Annals of Statistics* 1983; *11*,86-94.
- [23] Pinheiro JC, Bates D. Mixed-Effects Models in S and S-Plus. New York: Springer. 2000
- [24] Dunn PK, Smyth GK. Randomised quantile residuals, *Journal of Computational and Graphical Statistics* 1996; *5*, 236-244.
- [25] Van der Linden, WJ, Hambleton RK. Handbook of Modern Item Response Theory. New York: Springer, 1997.
- [26] Fox, J-P. Bayesian Item Response Modeling. New York Springer, 2010.

ANALYSING COGNITIVE TEST DATA: DISTRIBUTIONS AND NON-PARAMETRIC RANDOM EFFECTS 23

MRC UNIT FOR LIFELONG HEALTH AND AGEING, LONDON, UK, PHONE NUMBER: +44(0)2076705719

*E-mail address:* `g.muniz@nshd.mrc.ac.uk`

DEPARTMENT OF STATISTICAL SCIENCE, UNIVERSITY COLLEGE LONDON, UK

STORM RESEARCH CENTRE, LONDON METROPOLITAN UNIVERSITY, UK

STORM RESEARCH CENTRE, LONDON METROPOLITAN UNIVERSITY, UK