Normalized Mini-Mental State Examination for Assessing Cognitive Change in Population-Based Brain Aging Studies

Viviane Philipps^{a, b} Hélène Amieva^{a, b} Sandrine Andrieu^{c–e} Carole Dufouil^{a, b}
Claudine Berr^{f, g} Jean-François Dartigues^{a, b} Hélène Jacqmin-Gadda^{a, b}
Cécile Proust-Lima^{a, b}

^aINSERM U897, ^bBordeaux University, Bordeaux, ^cUniversity of Toulouse 3, ^dDepartment of Epidemiology and Public Health, University Hospital of Toulouse, ^eINSERM UMR 1027, Toulouse, ^fINSERM U1061, and ^gUniversity of Montpellier 1, Montpellier, France

Key Words

Dementia · Cognition · Curvilinearity · Epidemiologic methods · Longitudinal studies · Mini-Mental State Examination

Abstract

Background: The Mini-Mental State Examination (MMSE) is widely used in population-based longitudinal studies to quantify cognitive change. However, its poor metrological properties, mainly ceiling/floor effects and varying sensitivity to change, have largely restricted its usefulness. We propose a normalizing transformation that corrects these properties, and makes possible the use of standard statistical methods to analyze change in MMSE scores. **Methods:** The normalizing transformation designed to correct at best the metrological properties of MMSE was estimated and validated on two population-based studies ($n = 4,889$, 20-year follow-up) by cross-validation. The transformation was also validated on two external studies with heterogeneous samples mixing normal and pathological aging, and samples including only demented subjects. **Results:** The normalizing transformation provided correct inference in contrast with mod-

els analyzing the change in crude MMSE that most often lead to biased estimates of risk factors and incorrect conclusions.

Conclusions: Cognitive change can be easily and properly assessed with the normalized MMSE using standard statistical methods such as linear (mixed) models.

© 2014 S. Karger AG, Basel

Introduction

The Mini-Mental State Examination (MMSE) [1] is one of the most popular psychometric tests used to quantify global cognitive functioning and cognitive change in population-based longitudinal studies. MMSE consists of a series of questions aiming at quantifying global cognitive functioning on a 0–30 scale. In clinical practice, MMSE is widely used as a screening test or as part of the diagnosis of dementia [2], but it is also useful in patients' clinical follow-up to evaluate the severity of dementia, to decide whether or not initiating or stopping antidementia treatment, to estimate the prognosis, and to characterize the burden of dementia at a population level.

Yet, when exploring specifically cognitive change in aging populations mixing normal and cognitively impaired populations, the MMSE poses problems in statistical analyses because of its relatively poor metrological properties. The maximum MMSE score can be easily reached by a not cognitively impaired individual for whom the actual cognitive performance is thus not measurable [2]. This is particularly frequent among individuals with a high educational level (EL) [3]. Conversely, in severely impaired individuals, the minimum MMSE score may also be reached, making it impossible again to measure the actual cognitive performance. These well-known metric properties define, respectively, a ceiling effect and a floor effect. From a psychometric point of view, another limit strongly related to them is that a 1-point change in the score does not have the same clinical meaning according to the initial score. This nonstable sensitivity to change defines the curvilinearity. The MMSE has been shown to be a highly curvilinear psychometric test [4] since its sensitivity to change strongly varies with a very poor sensitivity to change in high scores (27–30) and a relatively good sensitivity to change in the medium range of scores (10–20).

Curvilinearity does not necessarily tarnish the discriminative ability of MMSE for dementia prediction. However, it implies that standard statistical modelling of predictors of MMSE scores is no more appropriate as it is mostly based on a gaussian assumption and a stable sensitivity to change. For instance, a previous paper showed that the linear mixed model used to describe change over time of MMSE and its predictors could conclude to spurious associations with risk factors [5]. To handle this curvilinearity, we previously developed a latent process model that was found very efficient to correct these biases [5, 6]. However, this model is rather complicated to implement and only available in specific software.

In this context, our objective was to provide and validate a simple normalizing transformation of the MMSE using the latent process model. Such normalizing transformation that would correct for the MMSE poor metrological properties aims at enabling the use of standard statistical methods to assess predictors of MMSE scores in population-based longitudinal studies.

Methods

Populations

Estimation and validation of the normalizing transformation of MMSE were based on data from two large population-based prospective cohorts of cognitive aging (PAQUID and Three-City

Study). The PAQUID study was established in 1988 to study cerebral aging and incidence of dementia [7]. Community-dwelling subjects lived in south-west of France and were at least 65 years old at the initial visit. They were followed up after 1 year and then every 2–3 years during 20 years. The multicenter Three-City (3C) study began in 1999 and aimed at evaluating the relationship between vascular factors and risk of dementia [8]. Subjects aged at least 65 years were recruited in 3 French cities: Bordeaux, Dijon and Montpellier. They were followed-up every 2–3 years during 10 years. For both cohorts, psychometric tests and dementia diagnosis based on DSM-III-R criteria were assessed at each visit.

From the original cohorts, MMSE scores at baseline were excluded due to a ‘first-passing effect’ previously evidenced [9]. This effect, possibly explained by the apprehension of the test situation at the inclusion visit, translates in an improvement between the two first visits. All the subjects (including subjects with dementia) who had at least one measure of MMSE from the first-year follow-up in PAQUID and the 2-year follow-up in 3C were included in the analysis.

Estimation and Validation Samples

In order to maximize the information, the PAQUID sample and the Bordeaux sample of the 3C study were pooled together in an ‘estimation sample’ to define the normalizing transformation. Validation of the normalizing transformation was done on external ‘validation samples’ that were 3C Montpellier and 3C Dijon samples. These samples also consisted in heterogeneous populations mixing normal and cognitively impaired populations. To further evaluate whether the transformation could also be applied in prodromal AD, we considered two subsamples constituted of all the subjects diagnosed with dementia (prevalent and incident) from the 3C Montpellier and 3C Dijon samples (called ‘demented 3C’) or from the PAQUID sample (called ‘demented PAQUID’).

Covariates

Numerous covariates were included in the statistical models described below: gender, EL defined in three classes (subjects who graduated from secondary school, subjects who graduated only from primary school, and subjects who did not graduate from primary school), age at baseline, a cohort indicator, and when available, an indicator of presence of one or two alleles $\epsilon 4$ of the Apolipoprotein E (ApoE4). Several timescales were also investigated: time from entry in the cohort, age (coded in decades from age 65) and time preceding and subsequent to the diagnosis of dementia in dementia samples.

Statistical Models

In studies of cognitive change in which cognitive ability is measured by MMSE, the metrological properties of MMSE can be corrected by applying a normalizing transformation of the score. This was proposed in the longitudinal setting with a latent process mixed model that simultaneously transformed the scores and modelled the transformed scores according to covariates in a linear mixed model [6, 10]. In this approach, the quantity of interest is the actual unobserved cognitive level that generated the scores. This actual cognitive level defines a latent process whose trajectory over time is explained by a standard linear mixed model including random effects to model the between-subjects variability and co-

Table 1. Description of the two samples (PAQUID, 3C Bordeaux) used to create the normalizing transformation of MMSE and of the four samples (3C Dijon, 3C Montpellier, Demented from PAQUID and Demented from 3C Dijon and Montpellier) used to validate the normalizing transformation of MMSE

	PAQUID	3C Bordeaux	3C Dijon	3C Montpellier	Demented 3C	Demented PAQUID
Period	1988–2009	1999–2010	1999–2008	1999–2010	1999–2010	1988–2009
Subjects	3,000	1,889	4,253	1,965	508	856
Observations	11,999	5,825	10,775	5,917	1,232	3,887
Observations per subject						
25% quantile	2	2	2	2	2	3
Median	4	4	3	3	2	4
75% quantile	6	4	3	4	3	6
Age at inclusion, years						
Mean	74.9	74.5	74.2	73.0	77.4	75.3
Standard deviation	6.6	5.0	5.5	5.4	5.8	6.2
Subjects at inclusion who						
Graduated only from primary school	1,337 (44.6%)	436 (23.1%)	588 (13.8%)	362 (18.4%)	107 (21.1%)	351 (41.0%)
Graduated from secondary school	672 (22.4%)	1,225 (64.8%)	3,374 (79.3%)	1,487 (75.7%)	327 (64.4%)	137 (16.0%)
Women at inclusion	1,764 (58.8%)	1,174 (62.1%)	2,631 (61.8%)	1,149 (58.5%)	297 (58.5%)	600 (70.1%)
Initial MMSE score						
Min	0	18	11	9	9	0
25% quantile	24	26	26	26	24	22
Median	27	28	28	28	26	26
75% quantile	28	29	29	29	28	28
Max	30	30	30	30	30	30
Subjects with						
Prevalent dementia at entry in the cohort	41 (1.37%)	50 (2.64%)	40 (0.94%)	42 (2.14%)	82 (16.1%)	41 (4.79%)
Dementia diagnosed at entry in the sample (1st- or 2nd-year follow-up)	17 (0.57%)	44 (2.33%)	52 (1.22%)	34 (1.73%)	86 (16.9%)	17 (1.99%)
Incident dementia during follow-up	798 (26.6%)	214 (11.33%)	251 (5.90%)	89 (4.53%)	340 (66.9%)	798 (93.22%)
Age at dementia diagnosis, years						
Mean	86.1	83.0	82.8	80.1	81.9	86.1
Standard deviation	5.9	5.7	5.7	6.2	6.0	5.9

variates to evaluate their impact on the cognitive trajectory. The link between the actual cognitive level and the observed scores consists in a parameterized transformation that normalizes the score by capturing its metrological properties (including the curvilinearity). Originally, all the parameters were estimated simultaneously, so that the estimated transformations could slightly differ from one regression structure to the other (and/or between populations), and parameter estimates of investigated risk factors (computed in the transformed scale) could not be quantitatively compared between studies.

Yet, as the normalizing transformation corrects curvilinearity, which is an intrinsic property of the psychometric test, the transformations should be relatively stable from one study to the other, and a unique transformation could be defined once for all. This work aimed at evaluating this assumption and proposing and validating a unique transformation for the MMSE. Indeed, based on such unique transformation, standard linear mixed models could be applied directly on the transformed data to analyze change in cognition measured by MMSE, and regression parameters could be quantitatively compared even on different populations as the transformed scale would remain the same.

Strategy of Analysis

In a first step, the stability of a normalizing transformation of the MMSE was investigated in a cross-validation analysis. For this, the estimation sample was repeatedly and randomly divided into two subsamples, the training and the test sets. The training set was used to estimate the transformation called H^* using a latent process mixed model called $M^{(\text{train})}$. The test set was used to assess the transformation H^* on independent data. Specifically, we estimated two models on the test set: one latent process mixed model called $M^{(\text{test})}$ in which the transformation was again estimated to fit at best the test data, and one linear mixed model called $M^{(\text{test})*}$ that was applied to the scores previously transformed by H^* . These two models had exactly the same regression structure. By regression structure, we mean covariate adjustment, timescale and shape of trajectory (quadratic or linear) with corresponding correlated random effects.

These steps of division and estimation were repeated 500 times, and for each repetition the two models $M^{(\text{test})}$ and $M^{(\text{test})*}$ were compared. Under our assumption of stability of the normalizing transformation, for a given regression structure, the results found by the linear model on transformed data [$M^{(\text{test})*}$] were expected to be in agreement with those found by the more flexible latent process model [$M^{(\text{test})}$]. Comparison of the regression parameters was done with percentages of variation between the estimates,

Fig. 1. Histograms of the MMSE distribution observed in PAQUID (a), 3C Bordeaux (b), 3C Dijon (c), 3C Montpellier (d) and the two subsamples Demented PAQUID (e) and Demented 3C (f). All the available repeated measures of MMSE were represented.

variances of the estimates (obtained by the delta method in the latent process model), and p values of the Wald tests for estimates of significance.

This cross-validation analysis was replicated several times by changing the structures of regression within $M^{(\text{train})}$ and $M^{(\text{test})}$ and between $M^{(\text{train})}$ and $M^{(\text{test})}$ in order to evaluate the stability of the transformations to very different settings.

In a second step, the unique transformation of MMSE was computed as the averaged transformed scores over repetitions, and was validated on different validation samples using the same indicators as for the cross-validation analysis.

Statistical analyses were conducted in R software V2.15.0 [11] with `lcmm` function of package `lcmm` V1.5.8 [12]. Statistical tests were computed at the 5% level of significance. Wald tests were used for statistical significance of parameters or combinations of parameters.

Normalizing Transformation

The normalizing transformation of MMSE preserves the original direction of the test with lower values indicating lower cognitive functioning. The transformation was rescaled in 0–100 (as $100 \cdot [H^*(Y) - H^*(0)]/[H^*(30) - H^*(0)]$), so that the minimal MMSE score 0 corresponds to a normalized MMSE score of 0 and the maximal MMSE score 30 corresponds to a normalized MMSE score of 100.

Results

Samples Description

The six samples used to create or validate the normalizing transformation are presented in table 1. The samples included 3,000, 1,889, 4,253 and 1,965 subjects in PAQUID, 3C Bordeaux, 3C Dijon and 3C Montpellier, respectively. Among them, 856 (28.5%) received a diagnosis of dementia in PAQUID, 308 (16.3%) in 3C Bordeaux, 343 (8.1%) in 3C Dijon and 165 (8.4%) in 3C Montpellier.

The mean age at inclusion varied little between cohorts from 73 (standard deviation, SD, = 5.4) years old in 3C Montpellier to 74.9 (SD = 6.6) years old in PAQUID. The percentage of women was also relatively stable between cohorts varying from 58.5% in 3C Montpellier to 62.1% in 3C Bordeaux. The median MMSE score at inclusion was 27 (interquartile range, IQR, 24–28) in PAQUID and 28 (IQR = [26, 29]) in the three other samples. The main difference was for EL: while 64.8–79.3% of subjects graduated from secondary school in the three 3C samples, only 22.4% graduated from secondary school in

PAQUID, which must be explained by the 10-year difference in the inclusion periods.

The ‘Demented 3C’ subsample was composed of 508 subjects with a mean age at dementia diagnosis of 81.9 (SD = 6.0); 58.5% were women and 64.4% graduated from secondary school. The median MMSE score at inclusion was 26 (IQR 24–28). The ‘Demented PAQUID’ subsample was composed of 856 subjects with a higher mean age at dementia diagnosis [86.1 (SD = 5.9) year], more women (70.1%) and only 16% who graduated from secondary school (57% graduated at least from primary school). The median MMSE score was the same at inclusion (26, IQR 22–28).

MMSE score distributions are displayed in figure 1. They are relatively similar in PAQUID, 3C Bordeaux, 3C Dijon and 3C Montpellier samples, i.e. very asymmetric with a strong ceiling effect. In the demented samples, MMSE scores are lower, but the distributions remain substantially asymmetric and still contain high MMSE scores. Indeed, these samples include only demented subjects but mix pre- and postdiagnosis scores.

Variability of the Normalizing Transformation

Various latent process mixed models were considered in order to appreciate the variability of the MMSE normalizing transformation. The transformation was systematically approximated by quadratic I-splines [10] with the same 7 nodes (0, 10, 20, 23, 26, 28, 30) chosen according to the MMSE distributions in PAQUID and 3C Bordeaux samples during the follow-up (fig. 1). Figure 2 displays these transformations estimated on the 6 samples for different regression structures. Exclusively in this figure (to make the comparison possible), the transformations were rescaled using intermediate MMSE values (20, 26, 29) for which we had enough observations in all our samples rather than extreme values (0 and 30) that were relatively rare in some samples, especially value 0. All the transformations have very close shapes whatever the sample and whatever the regression structure. Some differences may be seen in low values (probably due to the lack of information in this range of values) but remain moderate. The transformations exhibit a very low sensitivity to change in the highest MMSE scores and a higher sensitivity to change in intermediate scores, illustrating the curvilinearity issue. As an example, in the normalized scale, a 0.5 loss between levels of 0.8 and 0.3 corresponds to an observed loss of 2 points between MMSE 30 and 28, while the same loss in the normalized scale between 0 and –0.5 corresponds to an observed loss of around 6 points between 26 and 20 (i.e. three times greater).

Fig. 2. Normalizing transformations estimated in latent process mixed models with different regression structures on the PAQUID, 3C Bordeaux, 3C Dijon, 3C Montpellier samples and the two subsamples Demented PAQUID and Demented 3C. Exclusively in this figure, the transformed MMSE were rescaled using intermediate MMSE values 20, 26 and 29 ($H^*(y) = [H(y) - H(26)] / [H(29) - H(20)]$) so that the transformed MMSE does not range between 0 and 100.

The cross-validation (with 500 repetitions) was applied on the 4,889 subjects of PAQUID and 3C Bordeaux taken together, with a training set of 3,600 subjects. To estimate the transformation, four models [$M^{(train)}$] assuming a quadratic subject-specific trajectory with age and differing by the adjustment were estimated (no adjustment, adjustment for age at entry and/or cohort indicator, or adjustment also on education and gender). We chose a quadratic subject-specific trajectory with age to capture the accelerated cognitive decline in older ages previously found in PAQUID [6, 13]. For the models assessed on the test set [$M^{(test)}$], the timescale was either age or time since entry in the cohort, the subject-specific trajectory was linear or quadratic, and again different adjustments were assumed. All the cross-validation analyses showed similar results: the percentages of variation of the parameter estimates were always below 10%, and the variances estimated when using the previously obtained transformation H^* in $M^{(test)*}$ were very close to the ones obtained with the latent process mixed model $M^{(test)}$ (re-estimating the transformation). Finally, the significance tests agreed most of the time in over 95% of the cases.

Table 2. Results of the cross-validation analysis on 4,889 subjects (3,600 for training and 1,289 for testing)

	Estimations			Standard errors		Significance, %		
	$M^{(test)}$	$M^{(test)*}$	%Var	$M^{(test)}$	$M^{(test)*}$	$M^{(test)}$	$M^{(test)*}$	agreement
Specification 1 ^a								
Intercept	77.718	78.663	1.22	1.538	1.390			
Age	-10.057	-9.932	1.24	0.953	0.923	100.0	100.0	100.0
Age 0	9.919	10.288	3.73	1.805	1.746	100.0	100.0	100.0
Cohort	-2.515	-2.674	6.33	1.794	1.699	28.0	33.4	93.8
Age × age 0	-5.218	-5.361	2.75	0.913	0.893	100.0	100.0	100.0
Age × cohort	5.121	5.180	1.14	1.172	1.126	99.8	100.0	99.8
Specification 2 ^b								
Intercept	65.229	65.697	0.72	1.791	1.525			
Time	-2.258	-2.176	3.64	4.113	3.908	4.8	4.6	98.2
Time ²	-3.271	-3.316	1.36	2.407	2.289	27.2	28.6	91.4
Gender	-0.627	-0.621	0.86	1.108	1.099	7.0	7.0	99.2
EL-1st	12.375	12.226	1.21	1.477	1.412	100.0	100.0	100.0
EL-2nd	17.143	16.929	1.25	1.511	1.416	100.0	100.0	100.0
Age 0	-6.280	-6.207	1.16	0.949	0.923	100.0	100.0	100.0
Time × gender	3.111	3.091	0.63	2.985	2.943	14.4	15.8	96.6
Time × EL-1st	4.187	4.221	0.80	3.860	3.718	14.0	16.4	97.2
Time × EL-2nd	7.876	7.841	0.44	3.966	3.781	58.0	58.4	95.2
Time × age 0	-6.854	-6.821	0.49	2.797	2.754	72.2	72.4	95.0
Time ² × gender	-0.250	-0.243	2.94	1.843	1.828	2.6	3.6	98.6
Time ² × EL-1st	-2.819	-2.790	1.01	2.267	2.184	21.0	20.6	97.6
Time ² × EL-2nd	-3.218	-3.128	2.78	2.447	2.332	24.4	23.4	97.8
Time ² × age 0	-1.835	-1.929	5.01	1.911	1.894	14.4	16.0	97.6

Models $M^{(test)*}$ consist in linear mixed models estimated on the testing subsample using the MMSE scores pretransformed by H^* , while $M^{(test)}$ consists in latent process mixed models in which the transformation is simultaneously estimated on the testing subsample. H^* was obtained on the training subsample from $M^{(train)}$ assuming a cohort- and subject-specific quadratic trajectory with age. Given are the estimated parameters in $M^{(test)}$ and $M^{(test)*}$, the percentage of variation between these estimates (%Var), the standard errors, the percentage of parameter significance with Wald test at level 5%, and the percentage of agreement between the two significance tests. Two regression structures (detailed below) are investigated.

^a Specification 1 assumes a subject-specific linear trajectory with age (age) adjusted for age at inclusion (age 0), cohort (cohort)

and their interaction with age. Cohort indicates 3C Bordeaux vs. PAQUID in reference; age 0 and age are in decades and centered at 65 years old.

^b Specification 2 assumes a subject-specific quadratic trajectory with time from entry in the cohort (time and time²) adjusted for gender (gender), education level (EL-1st and EL-2nd), age at inclusion (age 0) and their interaction with time and time². Gender indicates men vs. women in reference; educational level is in three categories with subjects who graduated from primary school only (EL-1st) and subjects who graduated from secondary school (EL-2nd) vs. subjects who did not graduate from primary school in the reference category; age 0 is in decades and centered at 65 years old; time is in decades.

Table 2 presents two examples with $M^{(train)}$ including the cohort variable and age with a quadratic trend and $M^{(test)}$ including either a subject-specific linear trajectory according to age and adjusted for age at inclusion, cohort indicator and their interactions with age, or a subject-specific quadratic trajectory according to time from entry, adjusted for gender, EL, age at inclusion and their interaction with time and quadratic time. All the other cross-validation analyses are available on request.

Normalized MMSE

According to this, we defined the normalizing transformation of the MMSE as the pointwise mean transformation over the 500 repetitions in a model assuming a quadratic subject-specific trajectory over age and no adjustment for covariates. Figure 3a provides the correspondence table between the raw scores of MMSE and the final mean normalized scores of MMSE (that define the final transformation H^*). Figure 3b displays these 500

a

MMSE score	Transformed value		
	mean	2.5% quantile	97.5% quantile
0	0.00	0.00	0.00
1	2.91	2.66	3.18
2	5.48	5.06	5.95
3	7.76	7.22	8.34
4	9.77	9.15	10.45
5	11.57	10.86	12.34
6	13.19	12.40	14.03
7	14.67	13.83	15.57
8	16.05	15.15	16.99
9	17.37	16.40	18.36
10	18.68	17.63	19.76
11	20.01	18.88	21.15
12	21.38	20.21	22.56
13	22.83	21.67	23.99
14	24.39	23.19	25.53
15	26.07	24.90	27.17
16	27.91	26.77	28.97
17	29.93	28.82	30.98
18	32.17	31.08	33.19
19	34.64	33.60	35.64
20	37.37	36.37	38.35
21	40.40	39.39	41.36
22	43.74	42.75	44.67
23	47.40	46.45	48.31
24	51.44	50.54	52.25
25	55.98	55.15	56.75
26	61.18	60.44	61.87
27	67.25	66.65	67.89
28	74.61	74.13	75.12
29	84.32	84.02	84.63
30	100.00	100.00	100.00

Fig. 3. a Correspondence grid giving for each crude MMSE score the final normalized MMSE score (mean) and the 2.5 and 97.5% quantiles obtained with the cross-validation. **b** Plot of the final normalized MMSE scores (crosses) computed as the mean value over the 500 normalizing transformations (red lines; color refers to the online version only) obtained in the cross-validation study with a latent process mixed model including age as timescale, quadratic subject-specific trajectory and no adjustment.

transformations and the 31 final normalized values to emphasize the stability of the transformation. As an example, to correct metrological properties, crude scores of 20, 24 and 28 become, respectively, 37.37, 51.44 and 74.61. This transformation underlines the nonstable sensitivity of MMSE to change: a 4-point difference in crude MMSE represents an actual difference in the normalized MMSE score of roughly 14 points between MMSE scores of 20 and 24 and of more than 23 points between MMSE scores of 24 and 28.

External Validation

This normalizing transformation was validated on external data using 3C Montpellier and 3C Dijon samples, as well as the subsamples of demented subjects described in table 1. We compared results obtained when reesti-

imating the normalizing transformation, when using the one provided in figure 3 or when using a standard linear mixed model on crude scores (most frequently done in practice). Various structures of regression were tested on each sample. For clarity, in table 3 we only present the results from a single model for each validation sample. We did not always choose the same structure to illustrate the consistency of the results. Globally, the proposed normalizing transformation provided results close to those obtained when reestimating the transformation (that specifically corrects the metrological properties on the targeted sample), while the linear mixed model produced much more biased estimates. For example, in 3C Dijon, we investigated the effect of EL on the cognitive decline from entry in the cohort. Using the linear mixed model, the effect at baseline was underestimated while

Table 3. Validation of the MMSE normalizing transformation on samples 3C Dijon (n = 4,253), 3C Montpellier (n = 1,965), Demented PAQUID (n = 856) and Demented 3C (n = 508)

	Parameter estimates						% of variation	
	model M		model M*		model M ^{lin}		M*/M	M ^{lin} /M
	β	p value	β	p value	β	p value		
3C Dijon ^a								
Intercept	65.61		70.01		89.51		6.7	36.4
Time	-1.48	<0.001	-1.26	<0.001	-0.80	<0.001	14.4	46.0
EL-1st	12.05	<0.001	10.29	<0.001	5.26	<0.001	14.6	56.4
EL-2nd	16.55	<0.001	14.19	<0.001	6.75	<0.001	14.3	59.2
Gender	0.43	0.496	0.32	0.565	0.37	0.148	26.7	13.4
Age 0	-0.63	<0.001	-0.55	<0.001	-0.24	<0.001	12.6	62.2
Time × EL-1st	0.06	0.846	0.04	0.869	0.23	0.110	33.5	258.6
Time × EL-2nd	0.22	0.436	0.16	0.468	0.37	0.002	25.6	71.7
Time × gender	-0.40	0.002	-0.33	0.004	-0.18	0.003	17.2	56.2
Time × age 0	-0.05	<0.001	-0.04	0.001	-0.04	<0.001	19.8	19.8
3C Montpellier ^b								
Intercept	68.74		69.05		89.91		0.5	30.8
Age	3.66	<0.001	3.64	<0.001	0.96	<0.001	0.5	73.6
ApoE4	0.35	0.815	0.42	0.773	0.001	0.966	20.2	99.8
Age × ApoE4	-2.18	0.080	-2.20	0.070	-1.51	<0.001	0.8	30.7
Demented 3C ^c								
Intercept	48.84		44.34		72.58		9.2	48.6
T _{prediag}	-3.75	<0.001	-3.77	<0.001	-2.77	<0.001	0.4	26.1
T _{postdiag}	-4.19	<0.001	-3.65	<0.001	-5.13	<0.001	12.7	22.6
Age _{diag}	-0.19	0.056	-0.18	0.055	-0.15	0.073	4.2	18.4
EL-1st	1.91	0.334	1.91	0.311	1.58	0.337	0.2	17.4
EL-2nd	4.70	0.006	4.70	0.004	3.46	0.017	0.1	26.5
T _{prediag} × Age _{diag}	-0.002	0.960	0.003	0.931	-0.01	0.657	280.2	547.3
T _{prediag} × EL-1st	0.06	0.926	0.07	0.914	0.02	0.954	21.4	58.3
T _{prediag} × EL-2nd	-0.39	0.458	-0.49	0.366	-0.11	0.768	26.8	72.3
T _{postdiag} × Age _{diag}	-0.17	0.008	-0.16	0.012	-0.16	0.007	7.3	5.1
T _{postdiag} × EL-1st	3.57	0.009	3.24	0.014	4.19	0.001	9.1	17.6
T _{postdiag} × EL-2nd	1.85	0.117	1.79	0.117	2.01	0.068	3.4	8.5
Demented PAQUID ^c								
Intercept	38.86		36.04		61.87		7.3	59.2
T _{prediag}	-2.92	<0.001	-2.98	<0.001	-2.65	<0.001	2.0	9.4
T _{postdiag}	-2.81	<0.001	-2.51	<0.001	-5.06	<0.001	10.5	79.8
Age _{diag}	0.01	0.912	0.01	0.867	-0.02	0.846	45.5	315.5
EL-1st	7.31	<0.001	7.40	<0.001	7.38	<0.001	1.3	1.0
EL-2nd	14.64	<0.001	14.86	<0.001	14.86	<0.001	1.5	1.5
T _{prediag} × Age _{diag}	0.07	<0.001	0.08	<0.001	0.04	<0.001	5.8	38.9
T _{prediag} × EL-1st	-0.31	0.023	-0.40	0.005	0.12	0.348	27.1	139.5
T _{prediag} × EL-2nd	-0.14	0.431	-0.25	0.153	0.62	<0.001	84.0	557.7
T _{postdiag} × Age _{diag}	-0.13	<0.001	-0.13	<0.001	-0.14	0.011	2.9	7.6
T _{postdiag} × EL-1st	-1.40	0.001	-1.36	0.001	-1.53	0.015	2.7	9.4
T _{postdiag} × EL-2nd	-1.71	0.001	-1.83	<0.001	-1.10	0.175	7.4	35.3

Given are: the estimated parameters (β) and the associated p value, respectively, in the latent process mixed model (M) in which the normalizing transformation is estimated simultaneously, in the linear mixed model on normalized MMSE data (M*) and in the standard linear mixed model applied to crude MMSE data (M^{lin}), as well as the percentage of variation between the estimations of M and M* and of M and M^{lin}. Different structures of models were assumed for each cohort. Each time, correlated random effects on the intercept and the time variables were considered. ^a Time is the time from entry in the cohort in years, educational level is in three categories with subjects who graduated from primary school only (EL-1st) and subjects who graduated from secondary school (EL-2nd) vs. subjects who did not graduate from primary school in the reference category, gender gives men versus women in reference, age 0 is the age at entry in the cohort centered at 65 years given in decades. ^b Age is the current age centered at 65 years given in decades and ApoE4 codes the presence of at least one allele ϵ 4 of Apolipoprotein E vs. no allele. ^c T_{prediag} and T_{postdiag} are, respectively, the time in years preceding and subsequent to the time of diagnosis, Age_{diag} is the age at dementia diagnosis centered at 80 years and given in years.

the effect on the slope with time was overestimated, leading to incorrect conclusions regarding EL as a risk factor of cognitive decline ($p = 0.004$ for the three-category EL covariate). In contrast, the two models that either used the proposed normalizing transformation or estimated it, found similar estimates and conclusions, indicating no association between EL and cognitive change (with, respectively, $p = 0.634$ and $p = 0.565$ for the three-category EL covariate). In 3C Montpellier, the effect of ApoE4 was evaluated on the cognitive trajectory with age, and again, the conclusions diverged between the naïve linear mixed model and the models correcting the metrological properties. Finally, in the dementia samples, we studied the pre- and postdiagnosis declines according to EL and age at diagnosis. Again, diverging conclusions regarding the effect of EL on pre- and postdiagnosis slopes were found with the naïve linear mixed model in the PAQUID sample. In addition, the linear mixed model results were in favor of a postdementia decline twice larger than the predementia decline in the reference group composed of individuals diagnosed at 80 years without diploma (-5.13 points/year after vs. -2.77 points/year before diagnosis in 3C sample), whereas the other models correcting the metrological properties found relatively similar declines (e.g. -4.19 points/year after vs. -3.75 points/year before diagnosis in the 3C sample when estimating the transformation and -3.65 vs. -3.77 when using the proposed transformation). In addition to the correct inference, the transformed scores also provide covariate effects that can be compared between analyses. For example, with the same model estimated on the two demented samples, we showed that the mean cognitive level at diagnosis for subjects with no diploma (and diagnosed at 80 years) was 8 points lower in PAQUID than in 3C (36.04 vs. 44.34) but only 3 points lower ($36.04 + 7.40 = 43.44$ vs. $44.34 + 1.91 = 46.25$) for subjects who graduated from primary school and 2 points lower for those who graduated from secondary school ($36.04 + 14.86 = 50.90$ vs. $44.34 + 4.70 = 49.04$) because of a more pronounced effect of EL in PAQUID. Rates of change according to EL could be compared similarly.

Discussion

We proposed a normalizing transformation of the MMSE that corrects its weak metrological properties and makes possible the use of standard statistical methods for continuous variables that are linear (mixed) models to study MMSE scores as a dependent variable.

In publications on cognitive change, MMSE scores have been mostly analyzed using linear models, although crude MMSE scores did not satisfy the underlying assumptions (gaussian dependent variable and constant sensitivity to change). Yet, an extensive simulation study recently showed that the effects of risks factors on MMSE change over time could be largely impacted by this violation with type I errors for tests of risk factors up to 90% [5].

To correct the asymmetric distribution of MMSE scores, some authors proposed to use z-scores. However, such standardization does not correct the ceiling/floor effect or the varying sensitivity to change [14]. Others applied Tobit regressions [15], but they do not correct the tricky varying sensitivity to change. Rarely, a transformation, the square root of the number of errors, was considered, which considerably reduced the biases [9]. In this work, we proposed to move one step forward by (a) using a transformation that was specifically estimated to correct at best the metrological properties of the MMSE on a large brain-aging population-based study, and by (b) validating this transformation on several external datasets. This transformation is derived from a latent process mixed model that initially aimed at simultaneously correcting at best the metrological properties of a scale and estimating a regression model on the underlying normalized version of the scale [10]. However, such latent process mixed model remained a too complex statistical method and, as the normalizing transformation was estimated along with the effects of the investigated risk factors (computed in the transformed scale), these effects could not be quantitatively compared between studies. Using a unique normalizing transformation defined once for all, statistical analyses should be easy using standard methods, and effects of investigated risk factors should be quantitatively comparable between studies as illustrated in the demented external samples. We emphasize that the proposed transformation is dedicated to the statistical analysis of MMSE as a dependent variable. By changing the interval between two successive values, we counteract the metrological problems with MMSE and provide a score that uses the same information as MMSE but with corrected intervals between successive scores. Since it preserves the rankings, the transformed score has exactly the same discriminative properties as crude MMSE when used to predict dementia.

Validation of the normalizing transformation was done in two steps. First, in a cross-validation study, we found that using a linear mixed model on transformed data by a previously defined normalizing transformation

was an appropriate alternative to the more complex latent process mixed model when the population of interest was close to the population used for defining the transformation. This was also appropriate when the two regression structures (used for defining and for validating the transformation) differed greatly. Although we investigated here different specifications of models for studying cognitive decline, we did not aim at comparing their epidemiological value. We considered a wide range of regression structures to underline the stability of the proposed transformation to different settings. In an external validation, we showed that the normalizing transformation of MMSE also applied to cohorts not used for the estimation, even when the validation cohorts differed substantially from those used for the estimation (cohorts including only prevalent and incident cases of dementia for instance).

Nevertheless, such transformation is not universal. As the transformation is directly linked to the distribution of the MMSE scores, the transformation applies only to cohorts in which an asymmetry in the MMSE scores is observed. We intended to validate the normalizing transformation on a cohort of initially demented subjects [16]. However, the MMSE distribution had already a gaussian shape with most values inside the 10–25 range. On these data, curvilinearity was no more an issue as the main curvilinearity occurs above 25. In contrast, such asymmetry always occurs in prospective population-based cohorts including very heterogeneous ages and mixtures of subjects with a normal aging and subjects at the preclinical phase of dementia, or in clinical studies focusing on the progression from mild cognitive impairment to dementia. In these cohorts, central with the current effort for prevention and for evaluation of interventions in prodromal AD, correcting the metrological properties of MMSE is crucial to correctly ana-

lyze cognitive functioning and properly evaluate determinants of cognitive decline. The normalizing transformation we proposed will make possible a correct analysis of MMSE scores and will permit a direct quantitative comparison of risk factor effects obtained in different longitudinal studies. Although developed and validated on the French-language version, it should apply to any valid translation of MMSE. Indeed, the asymmetric distribution of MMSE in heterogeneous populations is observed whatever the language used for administering the MMSE [3, 17].

The transformed scores are obtained from figure 3 or by using the R package NormPsy that computes the normalized transformation and provides a function for back-computing predictions in the crude MMSE scale.

Acknowledgments

This work was carried out within the MOBIDYQ project (grant 2010 PRSP 006 01), which is funded by the Agence Nationale de la Recherche. We thank the REAL.FR/DSA group for providing data from the REAL.FR cohort. These data have not been directly used in this paper but have contributed to this work development. The PAQUID study was funded by SCOR insurance, Agrica, Conseil régional of Aquitaine, Conseils Généraux of Gironde and Dordogne, Caisse Nationale de Solidarité pour l'Autonomie, IPSSEN, Mutualité Sociale Agricole, and Novartis Pharma (France). The Three-City Study was funded by Sanofi-Synthelabo, Fondation pour la Recherche Médicale, Caisse Nationale d'Assurance Maladie des Travailleurs Salariés, Direction Générale de la Santé, Conseils Régionaux of Aquitaine and Bourgogne, and Fondation Plan Alzheimer.

Disclosure Statement

There are no conflicts of interest.

References

- 1 Folstein MF, Folstein SE, McHugh PR: 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198.
- 2 Tombaugh TN, McIntyre NJ: The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc* 1992;40:922–935.
- 3 Launer LJ, Dinkgreve MA, Jonker C, Hooijer C, Lindeboom J: Are age and education independent correlates of the Mini-Mental State Exam performance of community-dwelling elderly? *J Gerontol* 1993;48:271–277.
- 4 Proust-Lima C, Amieva H, Dartigues JF, Jacqmin-Gadda H: Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *Am J Epidemiol* 2007;165:344–350.
- 5 Proust-Lima C, Dartigues JF, Jacqmin-Gadda H: Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *Am J Epidemiol* 2011;174:1077–1088.
- 6 Proust C, Jacqmin-Gadda H, Taylor JM, Ganiayre J, Commenges D: A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006;62:1014–1024.
- 7 Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P: Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *Int J Epidemiol* 1994;23:1256–1261.

- 8 Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* 2003;22:316–325.
- 9 Jacqmin-Gadda H, Fabrigoule C, Commenge D, Dartigues JF: A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *Am J Epidemiol* 1997;145:498–506.
- 10 Proust-Lima C, Amieva H, Jacqmin-Gadda H: Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *Br J Math Stat Psychol* 2013;66:470–487.
- 11 R Development Core Team R: A language and environment for statistical computing. 2012.
- 12 <http://CRAN.R-project.org/package=lcm>.
- 13 Proust-Lima C, Amieva H, Letenneur L, Orgogozo JM, Jacqmin-Gadda H, Dartigues JF: Gender and education impact on brain aging: a general cognitive factor approach. *Psychol Aging* 2008;23:608–620.
- 14 Morris MC, Evans DA, Hebert LE, Bienias JL: Methodological issues in the study of cognitive decline. *Am J Epidemiol* 1999;149:789–793.
- 15 Piccinin AM, Muniz-Terrera G, Clouston S, Reynolds CA, Thorvaldsson V, Deary IJ, Deeg DJ, Johansson B, Mackinnon A, Spiro A 3rd, Starr JM, Skoog I, Hofer SM: Coordinated analysis of age, sex, and education effects on change in MMSE scores. *J Gerontol B Psychol Sci Soc Sci* 2013;68:374–390.
- 16 Gillette-Guyonnet S, Nourhashemi F, Andrieu S, Cantet C, Micas M, Ousset PJ, Vellas B: The REAL.FR research program on Alzheimer's disease and its management: methods and preliminary results. *J Nutr Health Aging* 2003;7:91–96.
- 17 Matthews FE, Stephan BCM, Khaw K-T, Hayat S, Luben R, Bhaniani A, et al: Full-scale scores of the Mini Mental State Examination can be generated from an abbreviated version. *J Clin Epidemiol* 2011;64:1005–1013.