# Methods

## IRT assumptions

### Unidimensionality

Unidimensionality of the CA-DFI scale was evaluated with a Confirmatory Factor Analysis (CFA) of the inter-item polychoric correlation matrices using the R-package (version 3.2.3) *Lavaan* (version 0.5-22). The fit of CFA was assessed by examining multiple indices and recommended criteria: Comparative Fit Index (CFI>0.90), Tucker-Lewis Index (TLI>0.90) and Root Mean Square Error of Approximation (RMSEA<0.10) [[21](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_21),[22](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_22)]. Furthermore, an Exploratory Factor Analysis (EFA) was performed for which unidimensionality was reached when the first factor accounts for at least 20% of the variability and when the ratio of the variance explained by the first to the second factor was greater than 4 [[23](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_23)].

### Local independence

Local independence, closely related to unidimensionality, assumes that after controlling for the dominant factor, there should be no significant correlation among item responses [[23](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_23)]. To identify local dependence (LD) the residual correlation matrix achieved by the single factor CFA was examined. Possible LD was considered for residual correlation greater than 0.2.

### Monotonicity

Monotonicity assumes that item endorsements increase as the construct levels increases. A non-parametric IRT model was fitted to the data to evaluate monotonicity using Mokken scaling in the R-package *Mokken* (version 2.8.3). This model yields nonparametric IRT response curve estimates, shows the probabilities of endorsing response categories and can be visually inspected to evaluate monotonicity. Full scale or individual item scalability coefficient H are interpreted as follows: 0.30-0.40 = weak scale, 0.40-0.50 = medium scale, and >0.50 = strong scale. To assure that items were scored in the same order by all respondents at all levels of DFI, a strong requirement in measurement practice, invariant item ordering (IIO) was also evaluated with coefficient HT as follows: 0.30-0.40 = weak IIO, 0.40-0.50 = moderate IIO, and >0.50 = strong IIO [[24](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_24),[25](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_25)].

## IRT model

Once the IRT assumptions were confirmed, a 2-parameter logistic IRT model for polytomous response was performed using the R-package *mirt* (version 1.20.1) to estimate the item parameters: discrimination (or slope α) and category thresholds (or locations βn). A higher discrimination value means that the item is more powerful to distinguish individuals with proximate levels of DFI. Item thresholds indicate item difficulty and locate item modalities along the latent construct where the item best discriminates among individuals. The model also provides accurate latent scores – the level of DFI for each participant – estimated by taking into account each item properties (thresholds and discrimination). Thus, item thresholds and participants' latent scores are positioned on the same metric. Model fit was assessed using the S-X² statistic which compares the differences (residuals) between the observed and expected response frequencies. Items with an S-X² p-value of less than 0.001 were considered to have poor fit [[27](#_ENREF_27)].

## Differential item functioning

Differential item functioning (DIF) analyses were performed to examine whether the probabilities of responding in different categories vary across groups given equivalent levels of DFI [[27](#_ENREF_27)]. Two kinds of DIF were measured: non-uniform DIF which appears in the discrimination parameter and uniform DIF which refers to DIF in the threshold parameter. DIF was evaluated using the R-Package *Lordif* (version 0.3-3) using ordinal logistic regression models with a McFadden’s pseudo R² change of 2% as critical value [[26](file:///P:\2-%20PUBLICATIONS\Publi%2012%20-%20Epidemca_dep\Dementia%20&amp;%20Cogn%20Dis\Révisions\Epidep-DemCog_revised_%20Manuscript.docx#_ENREF_26)]. DIF was evaluated within the CA-DFI scale sample based on sex, age (< 75 years vs. ≥ 75 years), education level (≥7 years vs. <7 years of schooling), cognitive status (no disorder vs mild cognitive impairment (MCI) and dementia), living environment (urban vs. rural), country (*CAR* vs. *ROC*) and site (rural *CAR* vs. urban *CAR*, rural *ROC* and urban *ROC*).

## The area under the receiver-operating characteristic (ROC) curve

The area under the receiver-operating characteristic (ROC) curve (AUC) was calculated to determine the cutoff in the latent score that best differentiated the participants between dementia and dementia-free using the Youden index [[30](#_ENREF_30)]. ROC curve was evaluated using the R-package *pROC* (version 1.12.1) [[31](#_ENREF_31)].

The analyses were performed using SAS software (version 9.3; SAS Institute Inc., Cary, NC, USA) and R software (version 3.2.3 (2015-12-10); "Wooden Christmas-Tree" Copyright (C) 2015, the R Foundation for Statistical Computing).

# Evaluation of the IRT assumptions

## Unidimensionality

The CFA results indicated a good fit for a unidimensional model for the initial 16-item scale with CFI=0.988, TLI=0.986, and RMSEA=0.068. The first factor in EFA accounted for 59%, but the ratio of the variance explained by the first to the second factor was 3.5, which is under the published criterion of 4. Moreover, a Heywood case on the LEISURE item which presented a factor loading of 1.49 (> 1) was detected in the EFA analysis suggesting a poor confidence in the factors loadings. A step-by-step EFA procedure selection led to the removal of 6 items (LEISURE, PUBLIC, EDUC, MESSAGE, BATHING, and DRESSING). The EFA analysis conducted on the 10-item short form - including FIELDS, OCCUP, MARKET, FESTI, CEREMONI, FINANCES, MOBILITY, HOME, MEAL and LAUNDRY - produced the best criteria with a first factor at 73% and a ratio of 24.3. The indices resulting from the CFA analysis met the criterion with CFI=0.992, TLI=0.990 and RMSEA=0.082. Based on these results, the 10-item short form was considered as the best model. We thus assumed that the 10 items shared a single common factor and were sufficiently unidimensional to run in a 2-parameter IRT model*.*

## Local independence

No evidence of local dependence was found for the 10-item short form, whereas the residual correlation matrix of the 16-item form showed 4.2% of local dependent item pairs.

## Monotonicity

The Mokken scalability coefficient H of the 10-item form was 0.693, indicating a strong scale [[29](#_ENREF_29)] (vs. 0.634 for the 16-item form) with individual item scalability coefficients above 0.627 (see Table 2). The coefficient HT for IIO was 0.629. Inspecting response curve indicated no crossing between curves suggesting the item hierarchy was consistent over the latent construct. These results confirmed the assumption of monotonicity of the 10-item form and a strong IIO.

# Legend of supplementary figures

Supplementary Fig. A. DIF analyses for country, sites and living area. The EPIDEMCA program, Central Africa, 2011-2012 (n=301).

The overall impact of DIF for living area on the test characteristic curves (TCC). The TCC is the functional relation between the true score and the CA-DFI scale. A - Country (*CAR* vs. *ROC* for FESTI and LAUNDRY). B - Site (Nola vs. Bangui for FESTI and CEREMONI). C - Site (Nola vs. Gamboma for MOVE). D - Living area (rural vs. urban for FIELDS). The TCC shows the relation between the total item scores (y-axis) and the dependency construct (x-axis). The left graph show the TCC ignoring DIF for all the 10 items and the right graph shows the TCC for the item(s) having DIF.