

Sample size calculation and optimal design for univariate and multivariate regression-based norming

Francesco Innocenti, Frans Tan, Math Candel, & Gerard van Breukelen

Department of Methodology and Statistics
Maastricht University

ITC 2024 - Granada

- 1 Introduction
- 2 Research question
- 3 Optimal design
- 4 Sample size calculation
- 5 Extensions to multivariate norming
- 6 Current and future work

To prevent mistakes in the assessment of individuals, **norms should be precise**, that is, not being strongly affected by sampling error in the sample on which the norms are based.

How to **minimize sampling error** and **maximize precision** of the norms?

- 1 Adopt an efficient approach to norming \Rightarrow **continuous norming**
- 2 Find a sample composition (e.g. which age groups to include) that maximizes precision of estimation of the norms \Rightarrow the **optimal design**
- 3 Take a sufficiently large sample for the normative study \Rightarrow **sample size calculation formulas**

- **Inferential norming:** Angoff and Robertson [1987]; Zachary and Gorsuch [1985]; Zhu and Chen [2011]
- **Regression-based norming**
 - Multiple linear regression (MLR): Oosterhuis et al. [2016]; Van Breukelen and Vlaeyen [2005]; Van der Elst et al. [2011, 2005, 2006]
 - GAMLSS: Timmerman et al. [2021]; Voncken et al. [2019a,b]
- **Semi-parametric norming**
 - Quantile regression: Crompvoets et al. [2021]; Sherwood et al. [2015]; Vaughan et al. [2016]
 - cNORM: Gary et al. [2023]; Lenhard et al. [2019, 2018]

- ① Fit $\mathbf{y} = \mathbf{X}\beta + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$, thus obtaining $\hat{\beta}$ and $\hat{\sigma}$ from the normative sample
- ② To compare a **new individual** with the reference population:
 - Compute Z-score: $\hat{Z}_0 = \frac{Y_0 - \hat{Y}_0}{\hat{\sigma}} = \frac{Y_0 - \mathbf{x}_0^T \hat{\beta}}{\hat{\sigma}}$
 - Compute PR-score: $\hat{P}R_0 = \Phi(\hat{Z}_0) \times 100$

\mathbf{x}_0 = individual's scores on the predictors, $\Phi(.)$ = cdf of the standard normal distribution

- Simple and common approach (see delCacho Tena et al. [2024])
- **Limitations:** Normality & Homoscedasticity

- What is a design?

Joint distribution of the norm predictors in the sample given the sample size (N), e.g. sex distribution and age distribution per sex level in the sample

- What is the **Optimal Design** (OD)?

The joint distribution of the norm predictors in the sample that **minimizes** the sampling variance of the norm statistic (e.g. Z-score, PR-score) given N

- Innocenti et al. [023a]: **OD** is obtained by **minimizing the maximum of the sampling variance** of Z-score and PR-score over all possible combinations of the levels of the norm predictors, given N .

Optimal design: Results

Let $\epsilon_i \sim N(0, \sigma^2)$

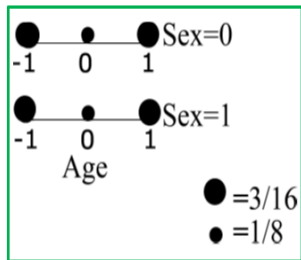
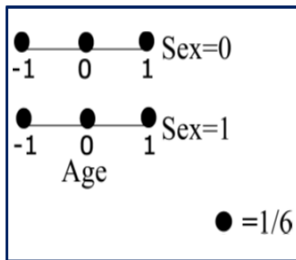
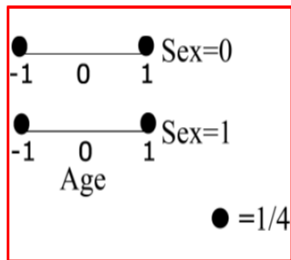
$$Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \epsilon_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Age}_i^2 + \epsilon_i \quad (2)$$

$$Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_4 \text{Age}_i \text{Sex}_i + \epsilon_i \quad (3)$$

$$Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Age}_i^2 + \beta_4 \text{Age}_i \text{Sex}_i + \epsilon_i \quad (4)$$

$$Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Age}_i^2 + \beta_4 \text{Age}_i \text{Sex}_i + \beta_5 \text{Age}_i^2 \text{Sex}_i + \epsilon_i \quad (5)$$

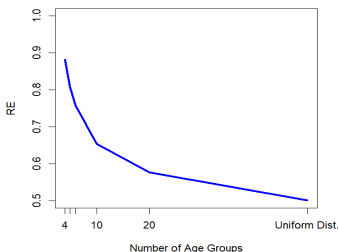


Why so few age groups under OD?

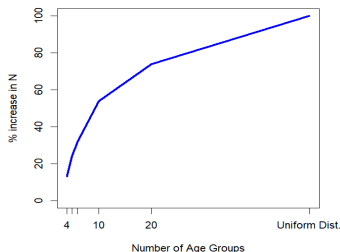
The 2/3 age groups required by OD follow from assuming a linear/quadratic age effect. **If this assumption is correct, including additional age groups yields a loss of statistical efficiency.** E.g.:

$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Age_i^2 + \epsilon_i$$

Relative Efficiency



% increase in sample size



Relative Efficiency (RE): ratio of sampling variance under OD to sampling variance under non-OD

% increase in sample size
relative to OD: $(RE^{-1} - 1) 100\%$

Sample size calculation for MLR-based norming

- Sample size requirements based on simulations in Oosterhuis et al. [2016], but limited to two norm predictors only
- Innocenti et al. [2023a] Sample size formulas for Z-score and PR-score **under OD** and any number of norm predictors
 - **Power:**
 - Norms application = **classification problem**, which can be expressed as: H_0 : "average" performance vs H_1 : "below average" performance given a chosen cut-off for classification
 - N^* = **to detect the smallest clinically relevant difference** between subject's norm value and the cut-off for classification, given pre-specified Type I error rate and statistical power
 - **Precision:** N^* = **half** the confidence interval **width equals** the pre-specified **margin of error**
 - Formulas based on delta method, which a simulation study has shown to be accurate for $N > 300$ for Z-scores and $N > 1600$ for PR-scores. Accurate = relative bias < 5%

Application (1/2)

1 Choose a norming model

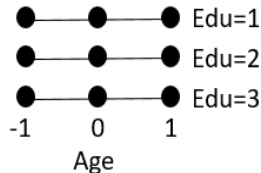
Van der Elst et al. (2006):

Letter M naming_i = $\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Low Edu} + \beta_4 \text{High Edu} + \epsilon_i$

$$\epsilon_i \sim N(0, \sigma^2)$$

2 Find the OD

9 age-education combinations of equal weight



3 Choose the norm statistic

Z-score = -1.64 (chosen cut-off for classification)

4 Sample size calculation

- **Power:** $H_0 : Z = -1.64$ vs $H_1 : Z < -1.64$, Effect Size (ES) = 0.36 (distance between 10th and 5th percentiles), $\alpha = 5\%$ and Power = 80%

$$N^* = \left[\frac{z_{1-\alpha} \left(k + 1 + \frac{Z_0^2}{2} \right)^{1/2} + z_{1-\beta} \left(k + 1 + \frac{(Z_0 - ES)^2}{2} \right)^{1/2}}{ES} \right]^2 = \left[\frac{1.64 \left(4 + 1 + \frac{1.64^2}{2} \right)^{1/2} + 0.84 \left(4 + 1 + \frac{(-1.64 - 0.36)^2}{2} \right)^{1/2}}{0.36} \right]^2 \approx 314$$

35 subjects for each age-education combination of OD

- **Precision:** confidence level $1 - \alpha = 0.95$, margin of error (MoE) = 0.18 (half distance between 10th and 5th percentiles)

$$N^* = \left[\frac{z_{1-\alpha/2} \left(k + 1 + \frac{Z_0^2}{2} \right)^{1/2}}{MoE} \right]^2 = \left[\frac{1.96 \left(4 + 1 + \frac{1.64^2}{2} \right)^{1/2}}{0.18} \right]^2 \approx 753$$

84 subjects for each age-education combination of OD

Sample size formulas implemented in R functions

Multivariate norming (1/3)

- Often normative studies derive norms for multiple tests with the same sample
- Univariate approach for each test is simpler but
 - Does **not take into account correlation** between test scores of the same subject -> **Incorrect classification** of subjects in clinical practice (see Agelink van Rentergem et al. [2019]; Su et al. [2015])
 - Multiple testing issues
- Current multivariate approaches
 - Van der Elst et al. [2017]: Same steps as MLR-based approach but using **multivariate multiple linear regression** -> What is the **multivariate performance** of a testee?
 - Agelink van Rentergem et al. [2018, 2019, 2017]
 - Advanced Neuropsychological Diagnostics Infrastructure (de Vent et al. [2016])
 - **Multilevel multivariate regression**
 - Multivariate performance summarized with **Hotelling's T^2** (Huizenga et al. [2007])

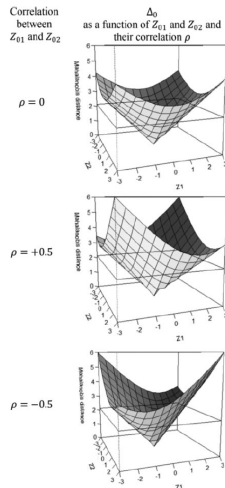
Multivariate norming (2/3)

- Innocenti et al. [023b]:

- 1 Multivariate multiple linear regression:
 $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, with $\mathbf{E} \sim N(\mathbf{0}, \Sigma)$
- 2 Multivariate performance summarized with Mahalanobis Distance (MD):

$$\hat{\Delta}_0 = \sqrt{(\mathbf{y}_0 - \hat{\mathbf{y}}_0)' (\hat{\Sigma})^{-1} (\mathbf{y}_0 - \hat{\mathbf{y}}_0)}$$

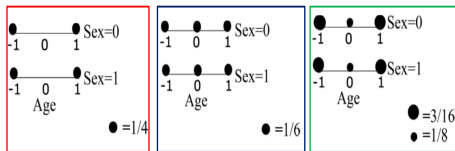
- MD = multivariate Z-score vs Hotelling's T^2 = multivariate t-statistic: Small differences in large samples, but MD made the derivation of OD easier
- Limitations** (shared with Van der Elst and Ageling van Rentergem's approaches):
Multivariate normality & homoscedasticity



Multivariate norming (3/3)

Sampling variance of MD similar to sampling variance of Z-score under univariate norming:

- **Same OD as the univariate case**



- **Sample size formulas similar to those for Z-score under the univariate case**, also implemented in R functions
- Sample size formulas based on delta method, which a simulation study has shown to be accurate if $N > 300$ and $MD > 1.18$ (median of χ^2 distribution with $df=2$). Accurate = relative bias < 5%, simulations limited to bivariate case.

with Dr. Alberto Cassese, University of Florence (IT)

- Sample size calculation for non-optimal designs
- Extensions of OD to models with 3 predictors (e.g. age, sex, education)
- Shiny Apps for sample size formulas
- Sample size calculation for interval estimation with assurance probability
- Simulation studies to assess accuracy of formulas for multivariate approach

- Sample size calculations and OD for most promising continuous norming approaches:
 - GAMLSS
 - cNORM
- Efficient designs that are robust against model misspecification at the design stage

References I

- Agelink van Rentergem, J., de Vent, N., Schmand, B., Murre, J., and Huizenga, H. (2018). Multivariate normative comparisons for neuropsychological assessment by a multilevel factor structure or multiple imputation approach. *Psychological Assessment*, 30:436–449.
- Agelink van Rentergem, J., de Vent, N., Schmand, B., Murre, J., and Huizenga, H. (2019). Predicting progression to parkinson's disease dementia using multivariate normative comparisons. *Journal of the International Neuropsychological Society*, 25:678–687.
- Agelink van Rentergem, J., Murre, J., and Huizenga, H. (2017). Multivariate normative comparisons using an aggregated database. *Plos One*, 12.
- Angoff, W. and Robertson, G. (1987). A procedure for standardizing individually administered tests, normed by age or grade level. *Applied Psychological Measurement*, 11:33–46.
- Crompvoets, E., Keuning, J., and Emons, W. (2021). Bias and precision of continuous norms obtained using quantile regression. *Assessment*, 28:1735–1750.
- de Vent, N., Agelink van Rentergem, J., Schmand, B., Murre, J., and Huizenga, H. (2016). Advanced neuropsychological diagnostics infrastructure (andi): A normative database created from control datasets. *Frontiers in Psychology*, 7.
- delCacho Tena, A., Christ, B., Arango-Lasprilla, J., Perrin, P., Rivera, D., and Olabarrieta-Landa, L. (2024). Normative data estimation in neuropsychological tests: A systematic review. *Archives of Clinical Neuropsychology*, 39:383–398.
- Gary, S., Lenhard, W., Lenhard, A., and Herzberg, D. (2023). A tutorial on automatic post-stratification and weighting in conventional and regression-based norming of psychometric tests. *Behavior Research Methods*, Advance online publication.
- Huizenga, H., Smeding, H., Grasman, R., and Schmand, B. (2007). Multivariate normative comparisons. *Neuropsychologia*, 45.
- Innocenti, F., Candel, M., Tan, F., and Van Breukelen, G. (2023b). Sample size calculation and optimal design for multivariate regression-based norming. *Journal of Educational and Behavioral Statistics*, Advance online publication.
- Innocenti, F., Tan, F., Candel, M., and Van Breukelen, G. (2023a). Sample size calculation and optimal design for regression-based norming of tests and questionnaires. *Psychological Methods*, 28:89–106.
- Lenhard, A., Lenhard, W., and Gary, S. (2019). Continuous norming of psychometric tests: A simulation study of parametric and semiparametric approaches. *PLoS ONE*, 14.

References II

- Lenhard, A., Lenhard, W., Suggate, S., and Segerer, R. (2018). A continuous solution to the norming problem. *Assessment*, 25:112–125.
- Oosterhuis, H., Van der Ark, L., and Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, 23:191–202.
- Sherwood, B., Zhou, A., Weintraub, S., and Wang, R. (2015). Using quantile regression to create baseline norms for neuropsychological tests. *Alzheimer's & dementia*, 2:12–18.
- Su, T., Schouten, J., Geurtsen, G., Wit, F., Stolte, I., Prins, M., Portegies, P., Caan, M., Reiss, P., Majoie, C., and Schmand, B. (2015). Multivariate normative comparison, a novel method for more reliably detecting cognitive impairment in hiv infection. *AIDS*, 29:547–557.
- Timmerman, M., Voncken, L., and Albers, C. (2021). A tutorial on regression-based norming of psychological tests with gamlss. *Psychological methods*, 26:357–373.
- Van Breukelen, G. and Vlaeyen, J. (2005). Norming clinical questionnaires with multiple regression: The pain cognition list. *Psychological Assessment*, 17:336–344.
- Van der Elst, W., Hurks, P., Wassenberg, R and Meijjs, C., and Jolles, J. (2011). Animal verbal fluency and design fluency in school-aged children: Effects of age, gender, and mean level of parental education, and regression-based normative data. *Journal of Clinical and Experimental Neuropsychology*, 33:1005–1015.
- Van der Elst, W., Molenberghs, G., Van Tetering, M., and Jolles, J. (2017). Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *The Clinical Neuropsychologist*, 31:1173–1187.
- Van der Elst, W., Van Boxtel, M., Van Breukelen, G., and Jolles, J. (2005). Rey's verbal learning test: Normative data for 1,855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. *Journal of the International Neuropsychological Society*, 11:290–302.
- Van der Elst, W., Van Boxtel, M., Van Breukelen, G., and Jolles, J. (2006). Normative data for the animal, profession and letter m naming verbal fluency tests for dutch speaking participants and the effects of age, education, and sex. *Journal of the International Neuropsychological Society*, 12:80–89.

References III

- Vaughan, R., Coen, R., Kenny, R., and Lawlor, B. (2016). Preservation of the semantic verbal fluency advantage in a large population-based sample: Normative data from the tilda study. *Journal of the International Neuropsychological Society*, 22:570–576.
- Voncken, L., Albers, C., and Timmerman, M. (2019a). Improving confidence intervals for normed test scores: Include uncertainty due to sampling variability. *Behavior Research Methods*, 51:826–839.
- Voncken, L., Albers, C., and Timmerman, M. (2019b). Model selection in continuous test norming with gamlss. *Assessment*, 26:1329–1346.
- Zachary, R. and Gorsuch, R. (1985). Continuous norming: Implications for the wais-r. *Journal of Clinical Psychology*, 41:86–94.
- Zhu, J. and Chen, H. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*, 29:570–580.

Thank you for your attention!

francesco.innocenti@maastrichtuniversity.nl

github.com/FInnocenti-Stat

researchgate.net/profile/Francesco-Innocenti