Sample Size Planning and Optimal Design for Estimating Regression-Based Reference Values

Francesco Innocenti, Frans E.S. Tan, Math J.J.M. Candel and Gerard J.P. van Breukelen

Department of Methodology and Statistics, Maastricht University

E-mail: francesco.innocenti@maastrichtuniversity.nl

ISCB 41 Krakow 2020

Introduction

- Normative studies are needed to obtain **reference values (or norms)** for **comparing patients** with the **reference population** in terms of relevant clinical measures, such as physiological variables (e.g. blood pressure, haemoglobin level), considering possible age or sex differences.
- In the traditional approach to norming, the sample drawn from the reference population is split into subgroups based on some demographic factors (e.g. age and sex), and then the norm statistics of interest are computed within each subgroup.
- The **regression-based norming approach**, instead, uses the whole sample and allows to identify the relevant predictors.
- To prevent mistakes in patients' assessment, **norms must be stable** (i.e. not being strongly affected by sampling error in the normative study).
- **Stability of norms** can be achieved by **carefully planning** both the **size** and the **design** of the sample from which the norms are derived (**normative sample**).

Regression-Based Norming

Denote by Y_i the outcome variable of interest (e.g. blood pressure), and by $\boldsymbol{x}_i$ the vector containing the demographic factors (e.g. age and sex) of patient *i*. A sample of *N* subjects is drawn from the reference population. The regression-based norming approach consists of the following steps:

1. Fit the **norming regression model:** $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$.
2. For patient *i*, compute the **predicted score:** $\hat{Y}_i = \boldsymbol{x}_i \hat{\boldsymbol{\beta}}$.
3. Compute the **residual error:** $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$.
4. Compute the standardized residual or **Z-score:** $\hat{Z}_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}}$.
5. If **normality** and **homoscedasticity** are met, the Z-score can be converted into a percentile rank score (PR) as follows

$$\widehat{PR}_i = \Phi\left(\hat{Z}_i\right) \times 100$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

Optimal Design

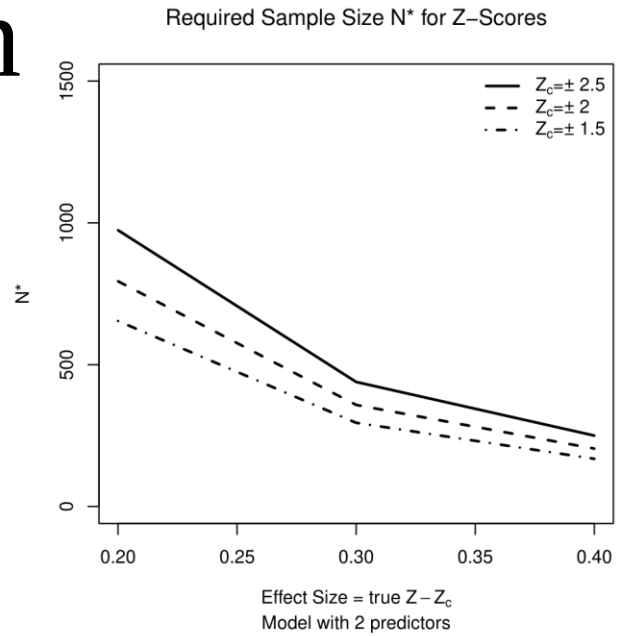
Some important definitions:

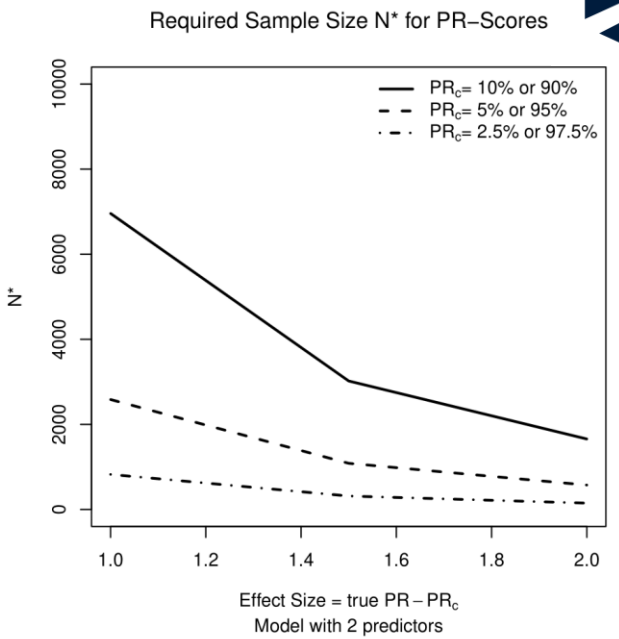
- **Design** = joint distribution of the predictors in the norming regression model, given *N*.
- **Support point** = a combination of the levels of the predictors.
- **Design weight** = proportion of the total *N* allocated to a support point.
- **Design region** = set of all possible support points.

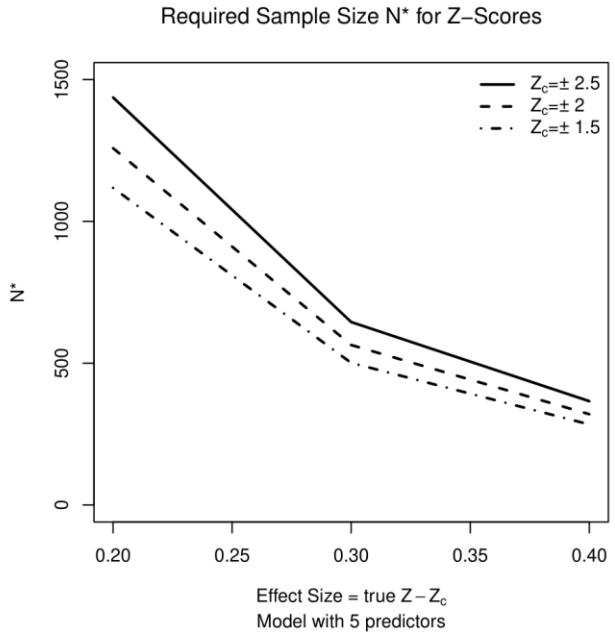
- The optimal design was derived for five regression models with a quantitative (e.g. age) and a qualitative predictor (e.g. sex), differing in whether they allow for interaction and nonlinearity.
- The **optimal design** was defined as **that joint distribution of age and sex that minimizes the sampling variances of** $\hat{Z}$ and $\widehat{PR}$ over the design region, given *N*.
- The optimal design
  - was **balanced** (i.e. same design weight for all support points) for models allowing either no or all possible interactions between age and sex, and **unbalanced** for models including some (but not all possible) interactions.
  - had **2 age levels** (at the boundary of the age range) for models allowing age to have only a linear effect, and **3 equidistant age levels** for models including a quadratic age effect.
- Since **the optimal design depends on the assumed model**, the most robust design against model misspecification was derived based on the relative efficiency and efficiency criteria. **The most robust design (i.e. highest minimum relative efficiency and highest minimum efficiency across all models) was balanced and had 3 equidistant age levels.**

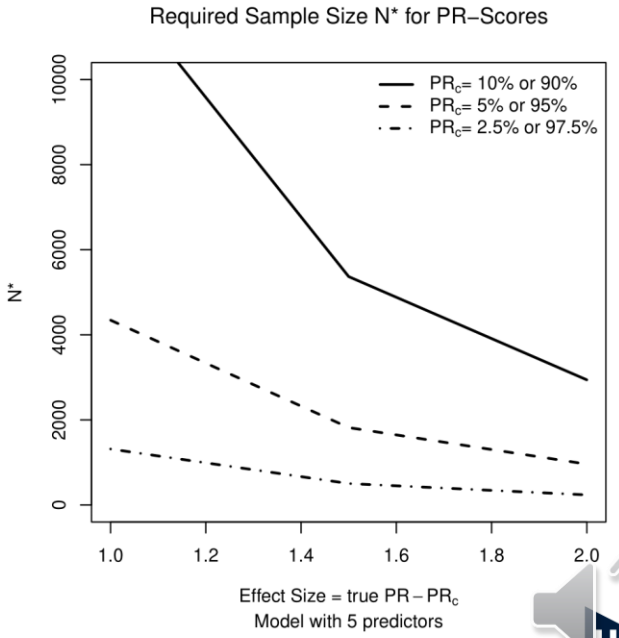
Sample Size Calculation

- Norms are used in practice to classify patient's symptoms relative to a chosen cut-off point as "normal" (e.g. $-2 < Z_i < 2$ or $5\% < PR_i < 95\%$) or "abnormal", to make decisions e.g. about clinical treatments.
- The **size** of the **normative sample must be sufficiently large** to allow **adequate classification** of the patient.
- A procedure was proposed to determine the size of the normative sample that allows to detect **the smallest "clinically relevant" difference between patient's true Z- or PR-score and the chosen cut-off point** (e.g. $Z_c = -2$ and $PR_c = 5\%$) used for decision making, given a pre-specified type I error rate and power.









Conclusion

- **For efficient estimation**, the **normative sample need not be representative** of the reference population with respect to the distribution of the predictors.
- The **stability of norms** can be **improved by** drawing normative samples as prescribed by the **optimal design under the assumed norming model**, or by the **most robust design in case of uncertainty about the underlying model**.
- The **required sample size depends on the norm statistic of interest**: Z-scores tend to require smaller sample sizes than PR-scores.