How could we design studies on PRO data?

The PLANIPRO project

Véronique Sébille, Myriam Blanchin, Alice Guilleux, Alexandra Rouquette, Sarah Amri, Mohand-Larbi Feddag, Tanguy Le Néel, Gildas Kubis, Bruno Falissard, Francis Guillemin, Jean-Benoit Hardouin

EA 4275 "Biostatistics, Pharmacoepidemiology and Subjective Measures in Health Sciences"
Université de Nantes
INSERM U 669, Universités de Paris 11 et Paris 5
EA 4360, Universités de Lorraine, Paris 5, Metz

ANR *N°2010 PRSP 00801*

UNIVERSITÉ DE NANTES

Background

- Evaluation of Patient Reported Outcomes (PRO)
  - How can we design these studies ?
  - Are studies adequately powered to determine clinically important changes in PRO ?
  - **Justification of study size is not always provided**

- WHY?
- Should we worry about it?

Importance of sample size

- *"Statistical analysis allows us to put limits on our uncertainty, but not to prove anything."*
  Altman DG. Practical Statistics for Medical Research. London, UK: Chapman & Hall; 1991.

- Clinical investigator's question: "How many individuals will I need to study?" …"It will only take 5 min"

- Adequate sample size likely to give enough power to detect a meaningful difference ⇨ ethical, clinical, methodological
  - Patients exposed to the burdens and risks of human research with a limited chance to provide any useful answers

Importance of sample size

- **Taking time to think about important issues**
  - The primary endpoint (secondary endpoints)
  - Expected clinically important difference on the primary endpoint
  - Type I and II errors
- Sample size needed for the *planning* and *interpretation* of clinical research

Sample size for PRO studies

- Clinical research methodology ⇨ has reached a high level of requirements
  - Publication of international guidelines (*CONSORT, STROBE, TREND, STARD, STREGA,* ***CONSORT PRO*** *…*)
    - "Study size"; “How sample size was determined"

- **What do (can, should?) we do for PRO studies?**
  - *Two main types of analytic strategies*
    - Classical test theory (**CTT**) ⇨ observed scores
    - Item Response Theory (**IRT**) ⇨ latent variable (latent trait)

Sample size for PRO studies

- **Classical test theory (CTT)** ⇨ observed scores
  - Most common framework
    - Sample size determination for normally distributed endpoints
    - Classical sample size formula

$$N = \frac{4\sigma_S^2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\delta_S^2}$$

⇨ Adequate power for CTT analyses

Sample size for PRO studies

- **Item Response Theory (IRT)** ⇨ latent trait
  - Assumed normally distributed
  - Most sample size calculations (*if any*) → Classical sample size formula for normally distributed endpoints
    - Inadequate for IRT → sample size underestimated

⇨ *BMC Medical Research Methodology, 2010;10:24.*

- **Consequences for sample size planning for IRT**
  - Latent *(≠ manifest)* variable + model → creates uncertainty on parameters

The PLANIPRO project

- **Main objective**
  - Provide valid sample size methodology
    - Comparison of PRO in two groups of patients (or between 2 times)
    - Cross-sectional & longitudinal studies
    - Using IRT modeling strategies *(Rasch and Partial Credit models)*

- **Proposed approach** ⇨ *Statistics in Medicine, 2012;31:1277-90.*
  - Analytical and numerical development based on the variance of the group (time) effect parameter & Wald test

Methods

- **Sample size**
  - Detect a group effect γ with power **1-β** and type I error α
    - Closely related to the Wald test of group effect
      - Based on an estimate Γ of γ and SE(Γ)
    - Derivation of SE(Γ)
      - Includes parameters related to the latent trait (means, variance), items of the questionnaire, sample size, expected patient's responses

Methods

- ***Planning phase of a study – Associated assumptions*** *(e.g. cross-sectional, Rasch model)*
  - Group effect $\gamma \rightarrow$ *expected* group effect (≥MCID)
  - Variance of the latent trait $\sigma^2 \rightarrow$ *expected* value
  - Number of items **J**, difficulty parameters δ_j (j=1,…,J)
    $\rightarrow$ *expected* values
  - Expected number of patients in each group **N/2**
    $\rightarrow$ linked to power for fixed α
  - Expected patients' responses $\mathbf{x_{nj}}$ (n=1,…,N) $\rightarrow$ *expected* responses / other *expected* parameters

Power of the test – The Raschpower method

- H_0: $\gamma = 0$ against H_1: $\gamma \neq 0$

Expected γ, δ_j, σ^2 and N_g

⇩

Expected **dataset** (patient's responses)

⇩

Estimation of γ and its **variance**

⇩

Estimation of the power **1-β_R**

Raschpower

Raschpower method – Does it work?

- Is Raschpower a **valid** approach for sample size planning for Rasch-family models?

- Is Raschpower **robust** to departures from the underlying modeling hypotheses?
  - Normality of the latent trait; local independence of items

- Investigated using simulation studies…

Simulation studies – Validity

- Simulated data → **Rasch model**
  - $\theta_0 \sim N(-\gamma/2, \sigma^2)$ et $\theta_1 \sim N(\gamma/2, \sigma^2)$
  - Variance of the latent trait $\sigma^2 \rightarrow 0.25, 1, 4, 9$
  - Group effect $\gamma \rightarrow$ *0.2 (small); 0.5 (medium); 0.8 (large)*

- Number of patients per group $N_0=N_1 \rightarrow$ *50, 100, 200, 300, and 500*

- Number of items **J** → *5 or 10*

- Difficulty parameters $\delta_j \rightarrow$ Normal or mixture of normal; possible gap Δ between the means of the latent trait and of the items parameters → $\Delta = 0, \sigma, 2\sigma$

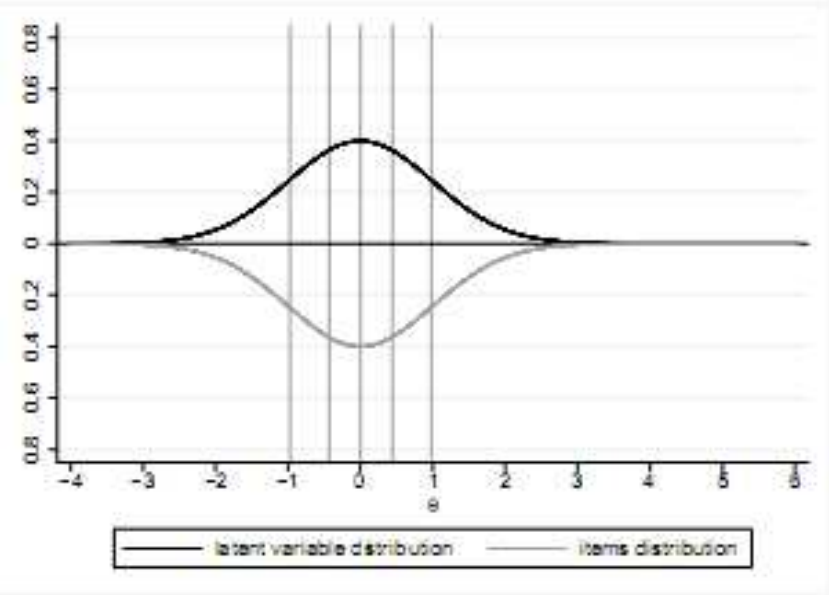
Methods – Simulated data

- **The Rasch model**
  - X_{nj}: response of patient n to item j
    - Realization x_{nj} (n=1,…,N; j=1,…J)
  - θ_n: realization of latent trait for patient n
  - δ_j: difficulty parameter for item j

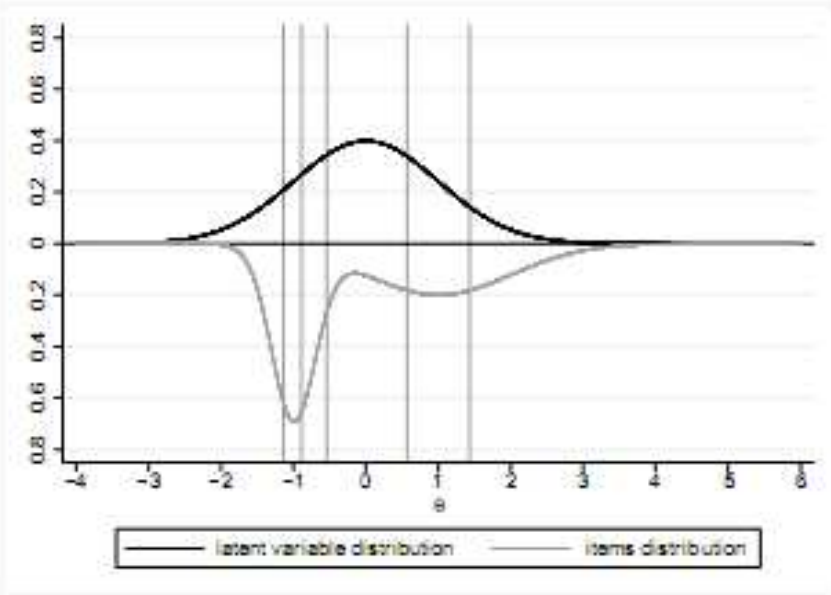
$$P\left(X_{nj} = x_{nj} \,\middle|\, \theta_n, \delta_j\right) = \frac{\exp\left\{x_{nj}\left(\theta_n - \delta_j\right)\right\}}{1 + \exp\left(\theta_n - \delta_j\right)}$$

  - $\theta_1, \theta_2, \ldots, \theta_N$ mutually independent, Gaussian distribution assumed

Items distributions – No gap $\Delta = 0$



(a) Normal distribution of items, $\Delta = 0$

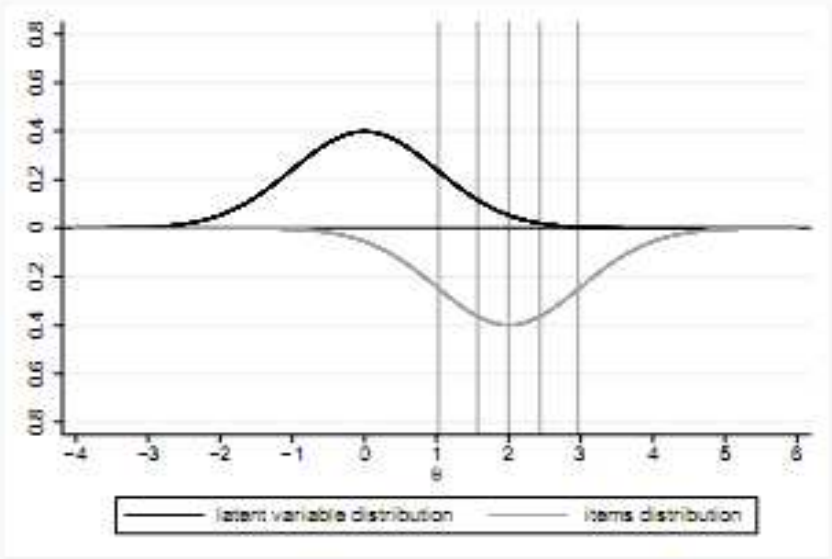


(b) Mixture of normal distributions of items, $\Delta = 0$

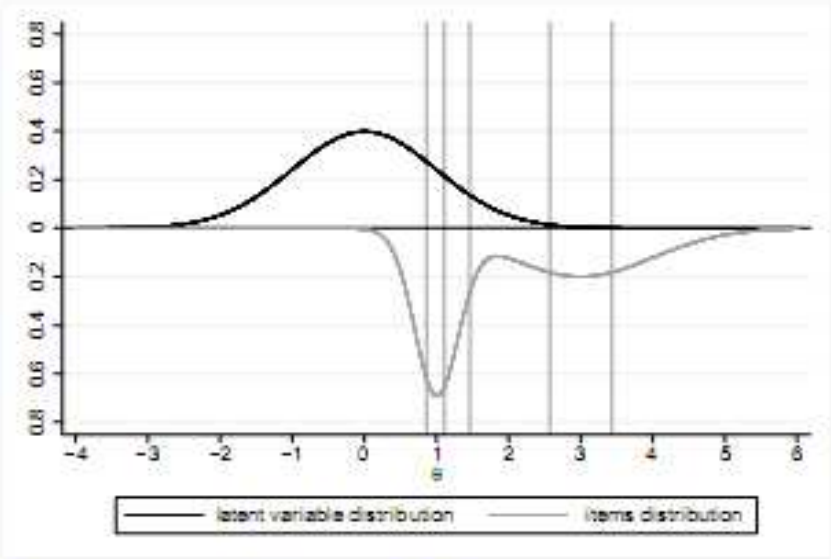
- *Regularly* spaced items difficulties
- Latent trait levels estimated with the *same accuracy* along the continuum

- *Irregularly* spaced items difficulties
- *≠ accuracy* of latent trait : e.g. more accurate around -1 / above -0.5

Items distributions – Gap $\Delta \neq 0$



(c) Normal distribution of items, $\Delta = 2\sigma$



(d) Mixture of normal distributions of items, $\Delta = 2\sigma$

- *Regularly* or *irregularly* spaced items difficulties
- *Gap* between distributions creates a *floor effect*: the most difficult items are too difficult for the population

Simulation study – Validity

- **For each replication *(simulation)***
  - Estimation of group effect + its variance → Mixed Rasch model including a group effect

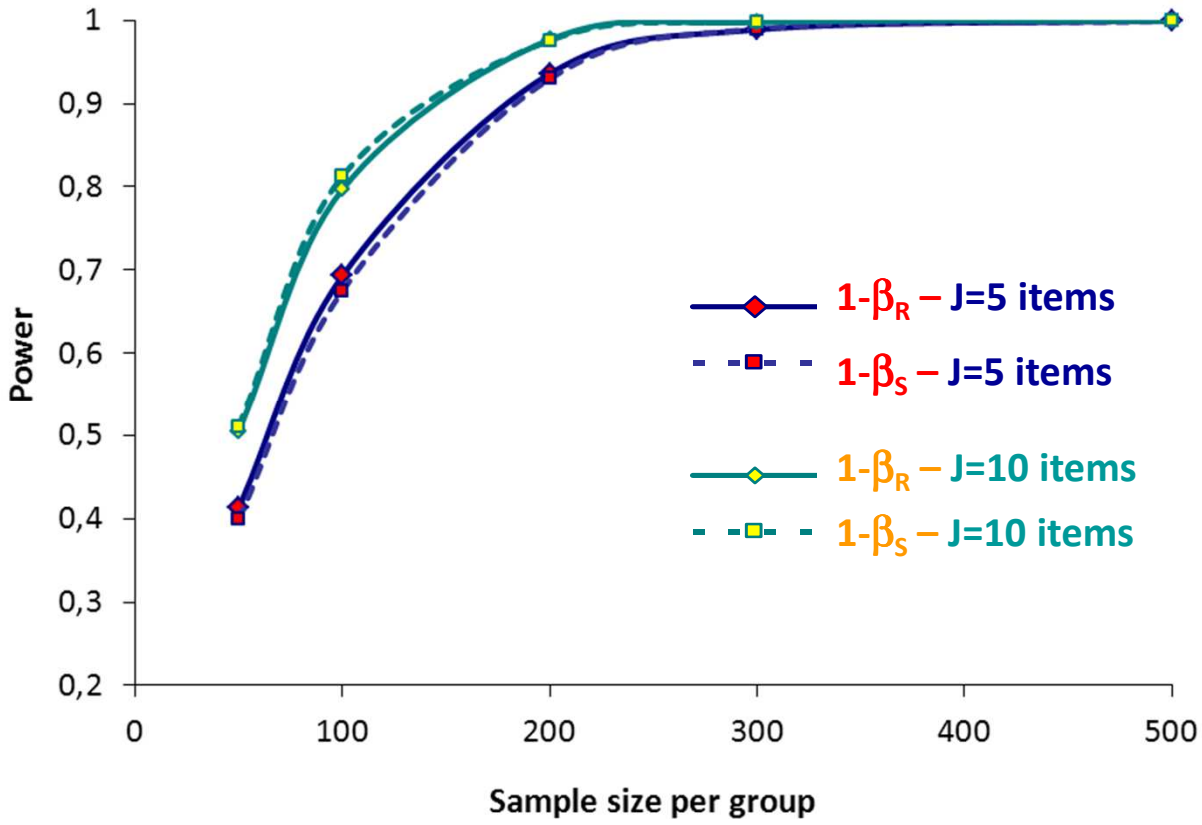
Difficulty items parameters
Variance of the latent trait

⇨ *Set to expected planning values*

- Wald test of group effect → estimated power $1-\hat{\beta}_S$
  - Rate of rejection of H_0 at $\alpha = 5\%$

⇨ 1,000 replications

Results – Power with Raschpower $(1-\beta_R)$, & simulations $(1-\beta_S)$ – γ=0.5; σ^2=1; δ_j Normal; Δ=0



For a given J

1- ≈ 1- S

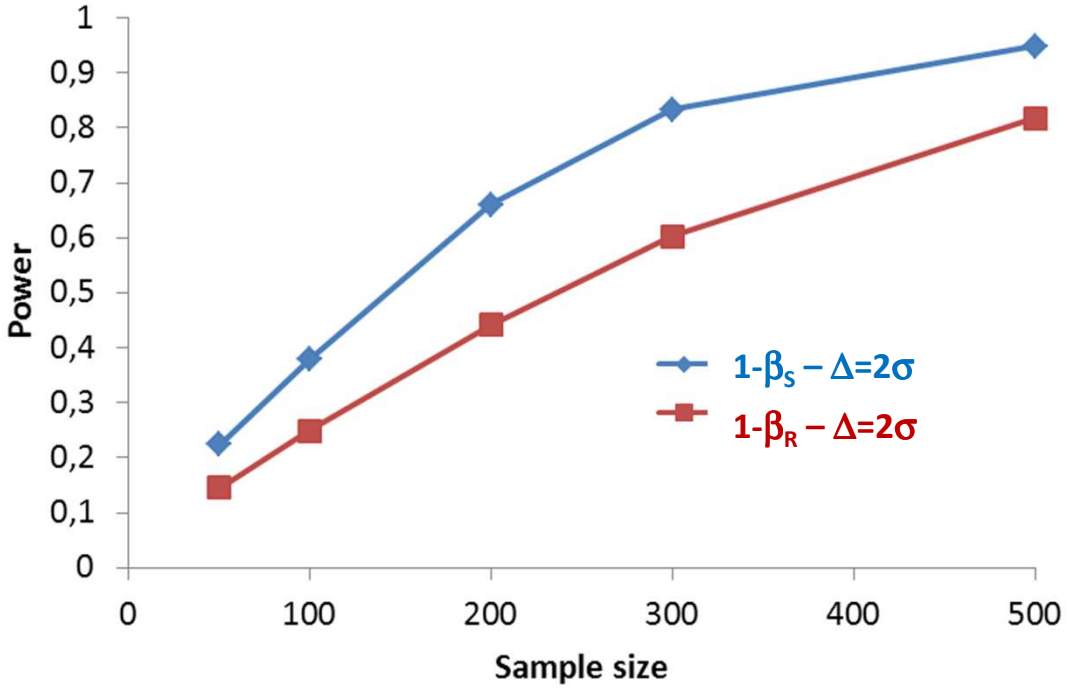
both ↑ with **J**

1- ≈ 1- S

For all values of σ^2 and all **items distributions**

Main results – Raschpower & simulations

- **Gap** between latent trait & items distributions *(γ=0.8; σ^2=9; J=5)*



Gap Δ=2σ

1- < 1- $_S$

Power underestimated with Raschpower

More marked as σ^2 and γ ↑

Raschpower method – Does it work?

- Raschpower seems to be a **valid** approach for sample size planning for Rasch-family models *(no gap)*
  - Cross-sectional studies, dichotomous and polytomous items *(data not shown here)*

    ⇨ *Plos One, 2013;8:e57279* ⇨ *Stat Med; under revision*

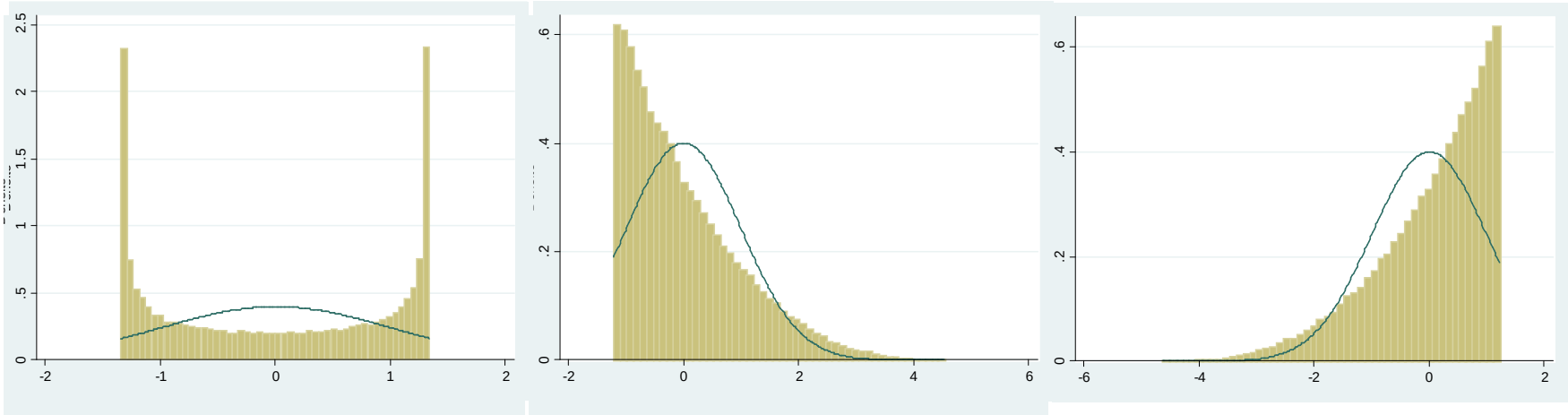
  - Longitudinal studies, dichotomous items *(data not shown)*

    ⇨ *J Appl Meas, 2014;in press.*

- **Gap** between latent trait & items distributions
  - Recommendation when planning studies: selecting the most appropriate questionnaire for the population
    - Avoid: using specific questionnaires for the general population

Simulation studies – Robustness

- The Raschpower method - Hypotheses
  - Normality of the latent trait
  - Locally independence of the items
- *What if* …. non-normal distribution of the latent trait



U shaped *L shaped* *J shaped*

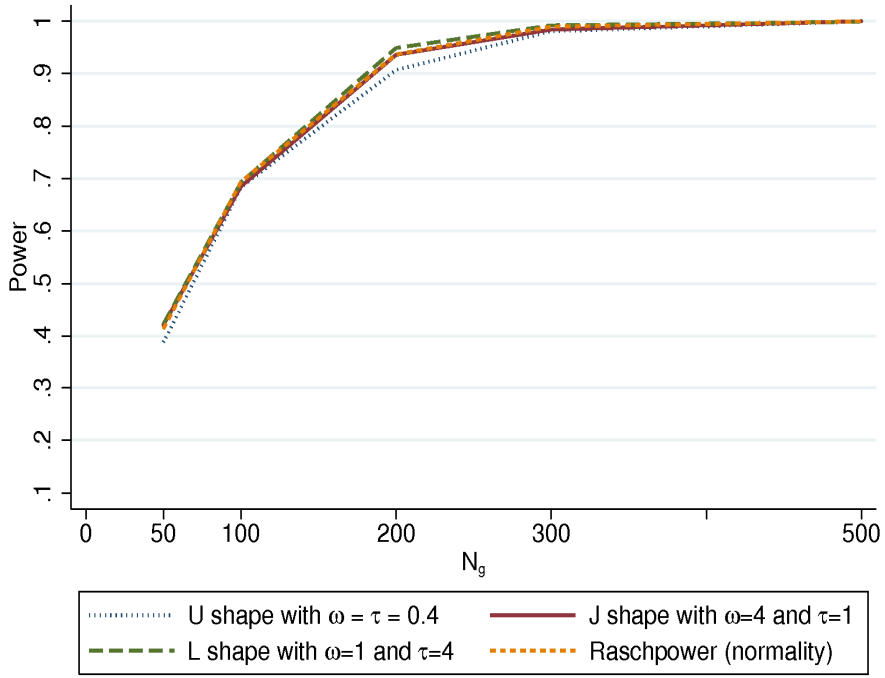
Simulated latent trait Simulated items difficulties

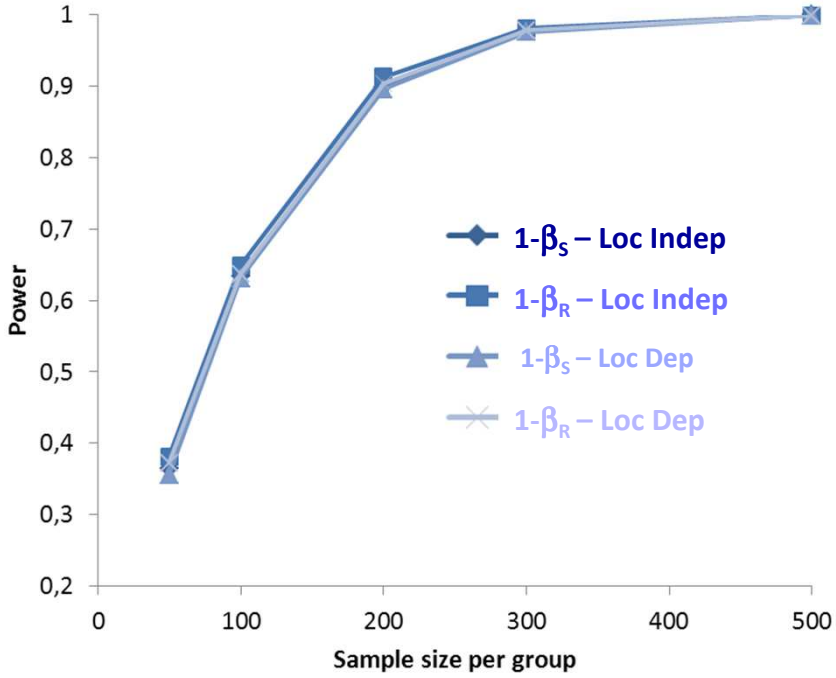
Robustness of Raschpower

- *What if* …. locally dependent items
  - e.g. SF-36 “Climbing one flight of stairs”; “Climbing several flights of stairs” etc.
- *Simulation* of dependent items (1 or 2 pairs of items)
- *Analyses*
  - Rasch model (assuming local independence)
  - IRT model *taking local dependence into account*
  - Raschpower (assuming local independence)
  - Raschpower *taking local dependence into account*

Robustness of Raschpower – Results

- **Power of test**





No impact on the power of the test

≈

Raschpower seems to be a **robust** approach for sample size planning for Rasch-family models *(cross-sectional studies, dichotomous items)*

⇨ Plos One, 2014;9:e83652 ⇨ manuscript submitted

Raschpower method – Planning made easy?

- Well…..planning of studies ⇨ many issues
  - A lot of assumptions regarding expected values of parameters
    - Variance of the latent trait (σ^2), items parameters (δ_j), group effect (γ)…
- What if we make wrong assumptions? What is the impact on Raschpower?
  - Misspecifications: σ^2 and items parameters δ_j
- Investigated using simulation studies…

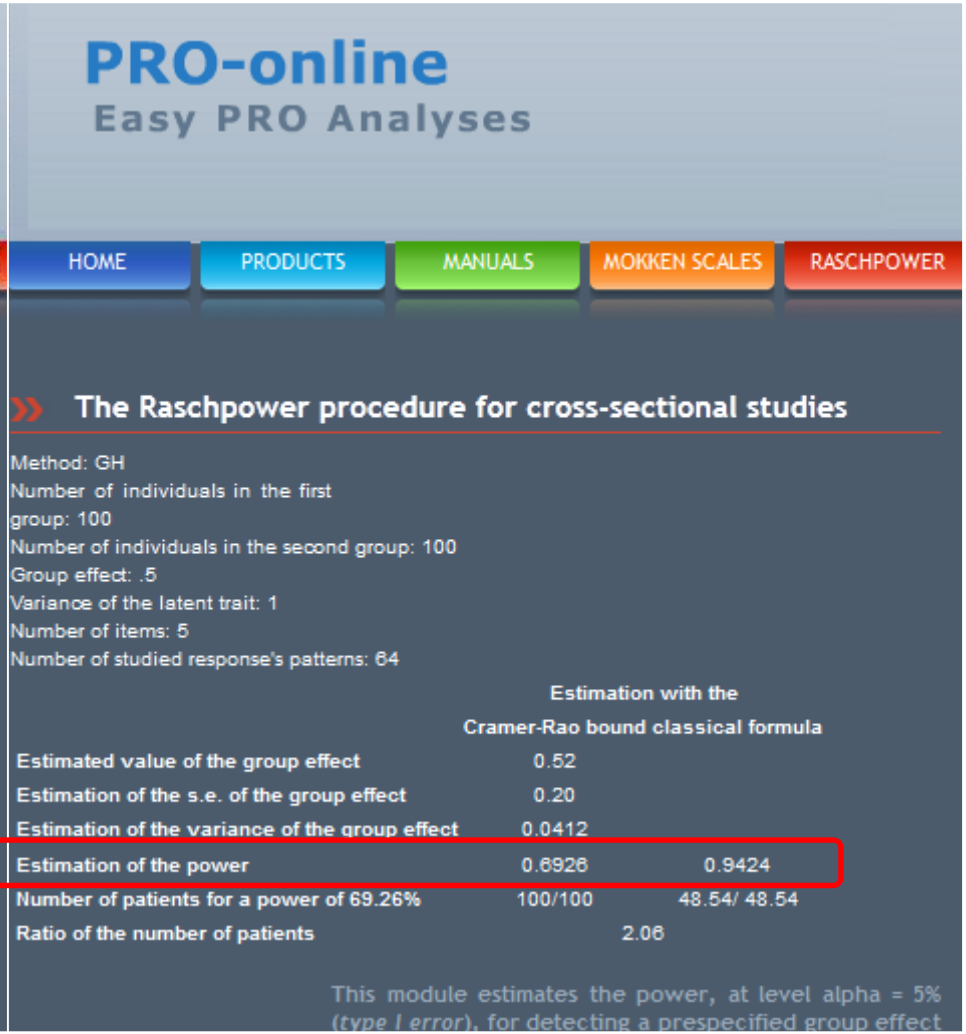
Misspecifications – Main results

- Misspecification of the variance of the latent trait
  - Underestimation of σ^2 ⇨ overestimation of $1-\beta_R$ ⇨ underpowered study
  - More impact if group effect $\gamma \geq 0.2$ and σ^2 small (<2)

- Misspecification of the items distribution
  - No impact on the power of the test of group effect given by Raschpower ($1-\beta_R$)

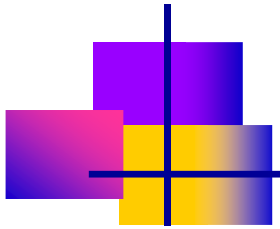
Raschpower in PRO-online (freely available)

http://pro-online.univ-nantes.fr



Discussion

- **Sample size / power calculations for the Rasch model**
  - Classical formula for manifest variables
    - Inadequate if Rasch model used for analysis
      - Underestimation of sample size
  - Development of the **Raschpower** method for power analysis
    - Seems valid and robust in ≠ situations
      - Cross-sectional / longitudinal studies
      - Dichotomous / polytomous items



Discussion

- **Sample size / power calculations for the Rasch model**
  - Two main parameters have an impact power
    - Size of the questionnaires (number of items J)
    - Heterogeneity of the sample (variance of the latent trait σ^2) ⇨ requires careful planning assumptions
      - Potential for sample size re-estimation
  - Importance of choosing suitable questionnaires for the population under study
    - Gap (between latent trait and items distributions) effect on power
      - BUT not specific to IRT

Discussion

- **Some drawbacks…**
  - Complexity of the approach?
    - Raschpower in PRO-online can help
    - Link between classical formula and Raschpower

      ⇨ *manuscript submitted*
  - Assumptions *(inherent to planning phase of studies)*
    - Underlying model
    - Size of group effect
    - Items parameters
    - Expected patient's responses (depending on previous assumptions)

Discussion

- **…and some perspectives**

  - The ≠ of the latent traits means γ ⇨ **interpretation** of a minimum clinically relevant ≠ on the latent trait scale? → unresolved issue yet…

    ⇨ *J Clin Epidemiol, 2014;67:433-40.*

  - How can we determine a MCID on the latent trait?
    ⇨ MIDIPRES project *(work in progress)*
    - **… and on the score? Have we reached consensus yet?**