Item response models for analysing assessments in rare diseases

Mats O Karlsson

Department of Pharmacy
Uppsala University
Uppsala, Sweden
Pharmacometrics
Pharmacokinetics - pharmacodynamics - disease progression

Plasma drug concentration
Neutrophil count
Tumor size (SLD)
Figure 1  Modeling and simulation (M&S) are performed before each decision point to quantitatively assess risk in moving forward. The drug and disease model is continuously updated to include new information acquired during drug development.
ADAS-cog test for Alzheimer’s Disease patients

Remembering Words

Naming Objects

Commands

Construction

Ideational Praxis

Orientation

ADAS-cog Score

Rosen et al. 1984
Types of Clinical Outcome Assessments

Patient-reported outcome (PRO) measures
Reports come directly from the patient

Clinician-reported outcome (ClinRO) measures
Reports come from a trained health-care professional using clinical judgment

Performance outcome (PerfO) measures
A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions

Observer-reported outcome (ObsRO) measures
Reports come from someone other than the patient or a health professional (e.g., a parent or caregiver) who has opportunity to observe the patient in everyday life
ADAS-Cog in Alzheimer’s Disease

Alzheimer's Disease Assessment Scale - Cognition

Tasks

Word-based

Rater assessed

Sum

0-70 score range
Responses assumed dependent on an underlying Cognitive Disability
Official IQ Test

The average IQ score in Europe is 100. Are you ready to test your brain?

→ Test My IQ Now!

How does it work?

1. 20 questions in 20 minutes
   Concentrate and answer our questions, which test your analytical memory, logic, numerical capabilities and

2. Immediate result
   Our algorithm estimates your IQ according to the official method and gives you your results as well as a scale to help you understand how you stand in relation to the population.

3. Get your certificate
   Get your certificate and performance report so you can share it with your family and friends.
Item Characteristic Curves ADAS-Cog

Reference:
Ueckert et al. Pharm Res 31(2014)
IRT Analysis of the SARA in ARCAs

Item Response Theory
Item-based analysis

Scale for the Assessment and Rating of Ataxia
Clinical Outcome Assessment (COA)

Autosomal Recessive Cerebellar Ataxias
Rare Neurodegenerative Disease (RND)
Autosomal Recessive Cerebellar Ataxias (ARCA)s
a heterogeneous group of rare and ultra-rare neurodegenerative diseases

Progressive disease
- Loss of coordination & ambulation

Affects the cerebellum and associated tracts

Genetically defined >200 disease types
How to measure the ataxia severity?
Scale for the Assessment and Rating of Ataxia (SARA)

- The most widely used outcome measure for ataxias
- Developed in 2004
- Clinician reported outcome (ClinRO)
SARA as a primary outcome measure in treatment trials?

**Problem**

- Concerns about SARA metric properties from regulatory agencies and recent studies
- Modifications to optimize the SARA
- Scarce data evidence and validation
- Analysis based on SARA total score

**Aim**

Evaluate the **metric properties and performance** of the SARA using **Item Response Theory (IRT)**
Sub-scores analysis using Item Response Theory (IRT)

IRT model of SARA

Latent variable \((\psi_i)\)

Ataxia severity

- Gait
- Stance
- Sitting
- Speech
- Finger chase
- Nose-finger
- Alternating hand
- Heel-shin

Observed item score \((Y_{ij})\)

\[ P(Y_{ij} = k) \]
Sub-scores analysis using Item Response Theory (IRT)

2-parameters logit functions

\[ P(Y_{ij} \geq k) = \frac{e^{(a_j(\psi_i - b_{j,k}))}}{1 + e^{(a_j(\psi_i - b_{j,k}))}} \]

\[ P(Y_{ij} = k) = P(Y_{ij} \geq k) - P(Y_{ij} \geq k + 1) \]

\( Y_{ij} \): observed item score for individual \( i \) and item \( j \)
\( k \): item response score

- **Scale characteristics**
  - \( a_j \): Item discrimination
  - \( b_{j,k} \): Item difficulty

- **Subjects characteristics**
  - \( \psi_i \): Latent variable

Item characteristic curves (ICCs)
Dataset
Autosomal Recessive Cerebellar Ataxias Registry

• 990 patients
• 1932 visits
• 69% of patients have genetically defined diagnosis
• 115 ARCA genetic subpopulations
• SARA sub-scores data
The questions we want to answer in this IRT analysis

- Do all SARA items share one common underlying latent variable?
- What are the characteristics (and performance) of each SARA item?
- Is one IRT model applicable to all ARCA genetic subpopulations?
The questions we want to answer in this IRT analysis, and how

Do all SARA items share one common underlying latent variable? (i.e., unidimensional)

Methods

- Data correlations
- Residuals correlations

What are the characteristics (and performance) of each SARA item?

- Item parameters
- Item characteristics curves
- Fisher information

Is one IRT model applicable to all ARCA genetic subpopulations?

- Model fit for each subpopulation
Item-pairs correlations to evaluate SARA dimensionality

Upper matrix
1. Data correlations → before modelling
2. Residual correlations → after modelling
3. Average correlations for 100 simulations

Lower matrix
4. Visual (VPC) diagnostic
   • The 5th, 50th, and 95th percentiles (lines)
   • 95% confidence intervals of the corresponding percentiles (shaded areas)
1. High (and similar) levels of correlations indicate unidimensionality
2. Data correlation patterns of simulated datasets mimic the original dataset
1. Low negative residual correlations indicate a good fit of the unidimensional model. 
2. Correlation patterns were mimicked in the simulations.
The questions we want to answer in this IRT analysis, and how

Do all SARA items share one common underlying latent variable?

What are the characteristics (and performance) of each SARA item?

Is one IRT model applicable to all ARCA genetic subpopulations?

Methods

- Data correlations
- Residuals correlations
- Item parameters
- Item characteristics curves
- Fisher information
- Model fit for each subpopulation
ICCs with separated curves and high discrimination parameters indicate properly designed response categories of SARA.
All SARA items are informative with varying importance at different disease severity levels.

Shaded areas: the ataxia severity interval for 95% of the studied population

\( I_j \): total item information in the population.
The questions we want to answer in this IRT analysis, and how

- Do all SARA items share one common underlying latent variable?
  - Methods
    - Data correlations
    - Residuals correlations

- What are the characteristics (and performance) of each SARA item?
  - Methods
    - Item parameters
    - Item characteristics curves
    - Fisher information

- Is one IRT model applicable to all ARCA genetic subpopulations?
  - Methods
    - Model fit for each subpopulation
### Analysis of ARCA genetic subpopulations

<table>
<thead>
<tr>
<th>Genetic diagnosis</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permuted group</td>
<td>996</td>
</tr>
<tr>
<td>Genetically undetermined</td>
<td>304</td>
</tr>
<tr>
<td>ARSACS</td>
<td>173</td>
</tr>
<tr>
<td>FA</td>
<td>110</td>
</tr>
<tr>
<td>One-subject diagnoses</td>
<td>69</td>
</tr>
<tr>
<td>SPG7</td>
<td>40</td>
</tr>
<tr>
<td>ATM</td>
<td>27</td>
</tr>
<tr>
<td>SYNE1</td>
<td>27</td>
</tr>
<tr>
<td>RFC1</td>
<td>25</td>
</tr>
<tr>
<td>SETX (AOA2)</td>
<td>25</td>
</tr>
<tr>
<td>TTPA (AVED)</td>
<td>21</td>
</tr>
<tr>
<td>ADCK3/COQ8A</td>
<td>18</td>
</tr>
<tr>
<td>POLG</td>
<td>16</td>
</tr>
<tr>
<td>ANO10</td>
<td>12</td>
</tr>
<tr>
<td>APTX (AOA1)</td>
<td>8</td>
</tr>
<tr>
<td>NPC1</td>
<td>8</td>
</tr>
<tr>
<td>PNPLA6</td>
<td>8</td>
</tr>
<tr>
<td>HEXA</td>
<td>8</td>
</tr>
<tr>
<td>ITPR1</td>
<td>7</td>
</tr>
<tr>
<td>CYP27 (CTX)</td>
<td>7</td>
</tr>
<tr>
<td>KIF1A</td>
<td>6</td>
</tr>
<tr>
<td>PEO1 (Twinkle)</td>
<td>5</td>
</tr>
<tr>
<td>PMM2</td>
<td>5</td>
</tr>
<tr>
<td>PNKP (AOA4)</td>
<td>5</td>
</tr>
<tr>
<td>CACNA1A</td>
<td>4</td>
</tr>
<tr>
<td>HARS</td>
<td>4</td>
</tr>
<tr>
<td>KCND3</td>
<td>3</td>
</tr>
<tr>
<td>KCNJ10</td>
<td>3</td>
</tr>
<tr>
<td>POLR3A</td>
<td>3</td>
</tr>
<tr>
<td>SIL1</td>
<td>3</td>
</tr>
<tr>
<td>STUB1</td>
<td>3</td>
</tr>
<tr>
<td>STXB1</td>
<td>3</td>
</tr>
</tbody>
</table>

→ Number of subjects in each subpopulation
Analysis of ARCA genetic subpopulations

Fit IRT model using all data

Model fit results: iOFVs

Random selection of 1 visit/subject

\[ \text{mean}(\text{iOFV}) \text{ for each subpopulation} \quad \text{mean}(\text{iOFV}) \text{ for entire ARCA population} \]

\[ \Delta \text{mean}(\text{iOFV}) \text{ and confidence intervals} \]

Fit better than expected    Fit worse than expected

\[ \Delta \text{mean}(\text{iOFV}) = \text{mean}(\text{iOFV}_{\text{subpop}}) - \text{mean}(\text{iOFV}_{\text{all pop}}) \]

iOFV: individual objective function value
Absence of evidence for differences between ARCA subpopulations

Permuted group:
a hypothetical subpopulation created by permuting the sub-scores of each item across individuals.

- Red points: difference in means of iOFVs
- Error bars: 95% confidence intervals (based on pooled two-sampled t-test assuming equal variances)
- n: number of subjects in each group

Results
- Data correlation
- Residuals correlation
- ICCs
- Fisher information
- Subpopulation analysis
Conclusions SARA Score

- Unidimensional- one single latent variable captures data
- SARA is well-performing with high discrimination values
- All items are informative with varying importance at different disease severity levels
- IRT model is applicable across genetic subtypes with no evident item patterns differences
Disease progression in the 10 ARCA subpopulations
Visual Predictive Check illustrate IRT model goodness of fit to total score data

- **Solid line**: the median of the observed data
- **Dashed lines**: 2.5th and 97.5th percentiles (dashed lines) of the observed data
- **Shaded areas**: 95% confidence intervals for the corresponding percentiles of the simulated data
- **Note**: total scores were calculated as the sum of the individual items score
VPC stratified on each item
Model parameter estimates and confidence intervals*

* Computed using sampling importance resampling (SIR) procedure

Base-slp IIV correlation: -0.5
**Clinical trial simulations and power/sample size calculations**

- **Simulate**
  - Full IRT model with hypothetical drug effect
  - Dataset with simulated sub-scores
    - 100 or 500 datasets

- **Estimate**
  - IRT model
    - Full model
    - Reduced model
  - IRT-informed TS model
    - Full model
    - Reduced model
  - Total score model
    - Full model
    - Reduced model

- **Both:**
  - Longitudinal data
  - End of treatment (base and last time point)

- **IRT power**
- **IRT-informed TS power**
- **TS power**

- Hypothesis testing:
  - Likelihood ratio test (LRT)
- Power-sample size curves:
  - Parametric power estimation* using SSE in PsN

*Ueckert et al., JPKPD 43:223-234 (2016)
IRT-informed link function for SARA scores

Total score

- $h()$ function is the disease progression function on the latent variable
- $\psi$: latent variable
- $Y$: total score
- $P_n$: polynomial functions describing the theoretical expectations of mean and variance. (derived from the base IRT model)

$$Y_{ij} = p_n(\psi_{ij}) + \varepsilon_{ij} \cdot p_n(\psi_{ij})$$
$$\varepsilon_{ij} \sim N(0, \sigma^2)$$
$$\eta_i \sim N(0, \omega^2)$$
**ARSACS - sample size for 80% power**

5-year study with disease-modifying treatment

Visits every 6 months

<table>
<thead>
<tr>
<th>Drug effect</th>
<th>Analysis</th>
<th>Sample size (treatment+control, 1:1) for 80% power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Early pop. 0-10 yr</td>
</tr>
<tr>
<td>50% inhibition</td>
<td>Tot score</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>IRT</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>IRT-inf</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>IRT EoT</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>IRT-inf EoT</td>
<td>302</td>
</tr>
<tr>
<td>100% inhibition</td>
<td>Tot score</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>IRT</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>IRT-inf</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>IRT EoT</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>IRT-inf EoT</td>
<td>94</td>
</tr>
</tbody>
</table>
100% inhibition

ARSACS - Comparison of sample sizes across different: Study durations & Analysis methods

![Graph showing sample size required for 80% power across different analysis methods and study durations.](image)
100% inhibition

POLG - Comparison of sample sizes across different: Study durations & Analysis methods

![Chart showing sample size required for 80% power across different analysis methods and study durations.]
Progressive Supranuclear Palsy (PSP)

• Neurodegenerative
  – Unknown cause
• Symptoms include:
  – Problems with walking, balance and eye movements
  – Cognitive impairment
  – Speech impairment
  – Difficulty swallowing
# Data description PSPRS

## Progressive Supranuclear Palsy Rating Scale

<table>
<thead>
<tr>
<th>Study</th>
<th>#patients</th>
<th>Mean #visits</th>
<th>Mean study duration (yr)</th>
<th>Mean disease duration (yr)</th>
<th>#arms</th>
<th>Mean age (yr)</th>
<th>Study type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbvie</td>
<td>377</td>
<td>4.3</td>
<td>1</td>
<td>3.2</td>
<td>3</td>
<td>69</td>
<td>interv</td>
</tr>
<tr>
<td>Biogen</td>
<td>161</td>
<td>2.8</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
<td>69</td>
<td>interv</td>
</tr>
<tr>
<td>Prospera</td>
<td>44</td>
<td>4.1</td>
<td>1</td>
<td>3.3</td>
<td>2</td>
<td>67</td>
<td>interv</td>
</tr>
<tr>
<td>Tauros</td>
<td>138</td>
<td>5.7</td>
<td>1.2</td>
<td>3.4</td>
<td>2</td>
<td>68</td>
<td>interv</td>
</tr>
<tr>
<td>Describe PSP</td>
<td>127</td>
<td>2.4</td>
<td>2.8</td>
<td>1.1</td>
<td>1</td>
<td>70</td>
<td>observ</td>
</tr>
<tr>
<td>ProPSP</td>
<td>132</td>
<td>2.6</td>
<td>1.6</td>
<td>4.3</td>
<td>1</td>
<td>69</td>
<td>observ</td>
</tr>
</tbody>
</table>
Item Characteristic Curves for PSPRS - Progressive Supranuclear Palsy Rating Scale
Item Characteristic Curves for PSPRS - Progressive Supranuclear Palsy Rating Scale

Star: FDA selected item
**PSP Ranking Scale items ranked by information content**

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Information content</th>
<th>Discrimination parameter</th>
<th>Number of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>2.16</td>
<td>2.97</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>1.99</td>
<td>2.86</td>
<td>5</td>
</tr>
<tr>
<td>26</td>
<td>1.91</td>
<td>2.93</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>1.4</td>
<td>2.29</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1.05</td>
<td>1.98</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>0.48</td>
<td>1.32</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>0.43</td>
<td>1.21</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>0.41</td>
<td>1.24</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0.39</td>
<td>1.14</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>0.35</td>
<td>1.1</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>0.34</td>
<td>1.19</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>0.34</td>
<td>1.08</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>0.33</td>
<td>1.03</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>0.29</td>
<td>1.07</td>
<td>5</td>
</tr>
<tr>
<td>21</td>
<td>0.28</td>
<td>1.17</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>0.26</td>
<td>0.96</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>0.24</td>
<td>0.93</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>0.21</td>
<td>1.01</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0.21</td>
<td>0.97</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>0.19</td>
<td>0.93</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>0.17</td>
<td>0.85</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>0.16</td>
<td>0.75</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>0.12</td>
<td>0.71</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>0.08</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>0.02</td>
<td>0.33</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>0.001</td>
<td>0.07</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.04</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>0.001</td>
<td>0.03</td>
<td>5</td>
</tr>
</tbody>
</table>

★ : FDA selected item
# Sample size calculation - PSP

- Parallel group study, 1:1 randomization placebo:active treatment
- 1 year study, 5 visits/patient, no dropout
- Treatment effect: 50% progression inhibition
- Disease progression rate as estimated from interventional studies
- 80% power

<table>
<thead>
<tr>
<th>scale</th>
<th>PSPRS</th>
<th>FDA</th>
<th>FDA rescore</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRT model (item level data)</td>
<td>36</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>IRT-informed model (total score data)</td>
<td>50</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>Linear mixed effects model (total score data)</td>
<td>62</td>
<td>76</td>
<td>84</td>
</tr>
</tbody>
</table>
Understand relation between endpoints

An IRT model can inform trial choices of population, instrument, design and analysis.
Separately determined ICCs

Example: Alzheimer’s Disease

- Utilize data from public or in-house clinical trial databases
- Study influence of patient population & assessment variant independent from another

Reference:
Ueckert et al. Pharm Res 31(2014)
Reduced tests options:

- Screening tests
- Trial conduct with limited tests
- Trial conduct with individualized dynamic testing

### Item Information

<table>
<thead>
<tr>
<th>Component</th>
<th>Information</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Delayed Word Recall</td>
<td>4.79</td>
<td>33.6</td>
</tr>
<tr>
<td>2 Word Recall</td>
<td>3.81</td>
<td>26.7</td>
</tr>
<tr>
<td>3 Orientation</td>
<td>1.64</td>
<td>11.5</td>
</tr>
<tr>
<td>4 Word Recognition</td>
<td>1.40</td>
<td>9.8</td>
</tr>
<tr>
<td>5 Naming O&amp;F</td>
<td>0.82</td>
<td>5.7</td>
</tr>
<tr>
<td>6 Number Cancellation</td>
<td>0.37</td>
<td>2.6</td>
</tr>
<tr>
<td>7 Construction</td>
<td>0.29</td>
<td>2.0</td>
</tr>
<tr>
<td>8 Word Finding</td>
<td>0.20</td>
<td>1.4</td>
</tr>
<tr>
<td>9 Ideational Praxis</td>
<td>0.18</td>
<td>1.3</td>
</tr>
<tr>
<td>10 Concentration</td>
<td>0.18</td>
<td>1.2</td>
</tr>
<tr>
<td>11 Remembering</td>
<td>0.16</td>
<td>1.1</td>
</tr>
<tr>
<td>12 Comprehension</td>
<td>0.16</td>
<td>1.1</td>
</tr>
<tr>
<td>13 Commands</td>
<td>0.15</td>
<td>1.1</td>
</tr>
<tr>
<td>14 Spoken Language</td>
<td>0.10</td>
<td>0.7</td>
</tr>
</tbody>
</table>

**90%**
Example: Parkinson’s Disease

+ Model links established (UPDRS) and novel endpoint (MDS-UPDRS)
  + Leverage historic data
  + Comparison with older compounds
  + Joint framework for complete disease severity range

+ Also done in AD for MMSE (often used for screening & diagnosis) & ADAS-cog (regulatory accepted endpoint)
  + Utilize all collected data
  + Leverage clinical routine data
  + Predict clinical endpoint from screening

References:
Gottipati et al. AAPSJ(2017)
Gottipati et al. PAGE 25 (2016) Abstr 5990
Jönsson et al PAGE (2017) Abstr 7236
**Biomarker – endpoint models**

**Multiple Sclerosis**

1\textsuperscript{st} generation model

- Natural History
- Drug Effect
- Disability

EDSS

2\textsuperscript{nd} generation model

- Cladribine
- ALC
- BoD

EDSS

**References:**
IRT model based subpopulation identification

- Parkinson ProgressionMarkers Initiative (PPMI) Database:
  - Subjects With Scans Without Evidence of Dopaminergic Deficit (SWEDD) (n = 64)
    - Consented as Parkinson’s patients
  - De Novo Parkinson’s Disease Subjects (n = 423)
    - Diagnosed ≤ 2 years
    - Not taking any medications for Parkinson’s disease
MDS-UPDRS total score

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeNovo</td>
<td>33</td>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>SWEDD</td>
<td>29</td>
<td>6</td>
<td>109</td>
</tr>
</tbody>
</table>
Item Response Model-based patient classification

Differentiation and prognosis of healthy subjects, SWEDDs and Parkinson's patients using a multi-dimensional item response theory model.
https://doi.org/10.1016/j.jns.2017.08.317
Note: In panel A, the concept within a circle is conceptualized as a latent variable; the smaller circles represent measurement error that contributes to the responses of each item; $\gamma$ denotes the causal effect of the concept on the item response. In panel B, the concept within a hexagon is conceptualized as a composite variable; $w$ indicates the weight (may or may not be equally weighted) used for the item response in computing the calculated composite score that represents the concept.
Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K)

<table>
<thead>
<tr>
<th>Condition</th>
<th>No 0</th>
<th>Yes +2</th>
<th>Yes +4</th>
<th>Yes +8</th>
</tr>
</thead>
<tbody>
<tr>
<td>New onset sensory or motor neuropathy involving cranial nerves</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lupus headache</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe, persistent headache (may be migrainous but must be nonresponsive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to narcotic analgesia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteinuria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;0.5 g/24 hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyuria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;5 WBC/high-power field; exclude infection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low complement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH50, C3, or C4 decreased below lower limit of normal for lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High DNA binding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased above normal range for lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBC &lt; 3 x 10^9/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exclude drug causes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0 points

Authors recommend a cutoff of 3 or 4 to define active disease and the need to increase therapy (see Evidence for details)
Potential of IRT modeling in clinical drug development

Plan
- Leverage more existing data (across compounds, populations, endpoints)
- Select specific patient populations
- Choose informative endpoints

POC

Analyze
- Infer with higher power
- Understand with increased detail

Plan
- Design more precisely (for regulatory accepted endpoint)
- Decide with increased confidence

Pharmacometric IRT Model
Potential of IRT modeling in clinical drug development

Phase III

- Inclusion criteria component
- Dynamic selection of tasks during trial

Enroll & run

- Futility analysis
- Adaptive design (drop arm, revision of sample size)

Interim analysis

- E-R analysis
  - Benefit-risk
  - Disease-modifying effect
  - Biomarker validation

Final analysis

Pharmacometric IRT Model
Acknowledgments

Evidence-RND consortium
- Matthias Synofzik
- Andreas Traschütz
- Rebecca Schüle
- Thomas Klockgether
- France Mentré
- Emmanuelle Comets
- Niels Hendrickx
- Ralf-Dieter Hilgers
- Nicole Maria Heussen
- Alex Sverdlov
- Yevgen Ryeznik
- Alzahra Hamdan
- Andrew Hooker
- Xiaomei Chen

Improve PSP consortium
- Günter Höglinger
- Carina Motz
- Martin Posch
- Franz König
- Elham Yousefi
- Mohamed Gewily

This work was funded by the European Joint Programme on Rare Diseases (EJP RD) Joint Transnational Call 2019 for the EJP RD WP20 Innovation Statistics consortia “EVIDENCE-RND” and “Improve PSP” focusing on “Innovative Statistical Methodologies to Improve Rare Diseases Clinical Trials in Limited Populations”.