Practical example: a computerized adaptive test for fatigue in rheumatoid arthritis

Stephanie Nikolaus, Christina Bode, Erik Taal, Harald Vonkeman, Cees Glas, Mart van de Laar
From theory to a “living” CAT
RA: Rheumatoid arthritis

Progressive, invalidating disease with changing course, chronic inflammation causes erosion and deformation of bones and joints

**Symptoms:** Swollen and painful joints, pain, stiffness, fatigue, restrictions in daily activities, worse mental and social well-being

**Clinical measures:** Laboratory values, radiographic damage, counting swollen and painful joints

**But:** Often no strong relation with subjective experience of health and quality of life = > **patient perspective:** pain, physical functioning, general health and fatigue

UNIVERSITY OF TWENTE.

ARTHRITE CENTRE TWENTE
Experience of fatigue in rheumatoid arthritis

- Annoying symptom with far-reaching consequences
- Multidimensional (fatigue is a physical experience, but also cognitive and emotional)
- Different from normal fatigue (no longer earned, more extreme)
- Fatigue is often neglected and has no role in the treatment of RA

“You feel like you’re carrying a load of flour on your back. That there’s always something weighing your shoulders down, that you’ve got to lug that around with you.” (woman, 58 years old)
Adequate measurement of fatigue in RA is important...

- ... for science and clinical practice
  - to get more insight into fatigue and its causes
  - for the evaluation of interventions or treatments to reduce fatigue

- ... for clinical practice
  - to facilitate communication about fatigue between patients and rheumatologists / health professional

- ... for society
  - fatigue is an important predictor of work absenteeism and participation restrictions in patients with RA
Existing fatigue questionnaires

One or more of the following disadvantages:

- Unsatisfactory content validity (not developed from patient perspective)
- Traditional format (fixed length-questionnaires)
- Patient burden possibly higher than necessary
- Not measuring everything that should be measured (not all dimensions included)
- Possibly only adequate for specific groups (e.g. highly fatigued patients)

→ Need for more appropriate and effective measurement
Solution: computerized adaptive testing (CAT)

- **Individual measurement** (per patient items selected from item bank based on previous response, tailored to individual level)

- **Precise and comprehensive** measurement, number of items to achieve same precision less as in traditional questionnaires

→ Use patient perspective and modern psychometrics

- Illustrate development of computerized adaptive test (CAT) for fatigue in RA and impression of first version $\text{CAT}_{\text{FatigueRA}}$
Preparation of the item pool

1) In-depth interviews on experience of fatigue

2) Q-sort study
   → Preliminary item pool consisting of items from existing questionnaires and new items based on interview material

3) Delphi study with experts to select adequate items
   → Content valid item pool (245 items among 12 dimensions)
Analysis of the item pool with IRT

4) Find out which of the fatigue dimensions and items, that came forward from previous studies, are adequate for the CAT development

- Examine how the items fit the underlying dimensions (IRT)
- Explore the dimensionality structure (factor analysis, Mplus)
  → application of the final multidimensional IRT model

Validated item bank to use for the construction of the CAT

Fatigue – a complex phenomenon

- Patients describe fatigue as multidimensional
- Three dimensions in $\text{CAT}_{\text{FatigueRA}}$ aiming to understand fatigue
  - Extra information gained by items of correlated dimensions can effectively be used in a CAT; leads to selection of most informative items
    - $\rightarrow$ greater measurement precision
Multidimensional item bank

- **Severity**: severity, duration, frequency (13 items)
  Example: Did you feel tired during the last 7 days?

- **Impact**: cognition/concentration, negative emotions, energy, sleep/rest, body feeling, coping, consequences (169 items)
  Examples: Have you felt down or dejected because of fatigue? / During the last 7 days, I could cope well with my fatigue. / How did your fatigue prevent you from cooking?

- **Variability**: change, perceived causes (14 items)
  Examples: How did your fatigue change during the last 7 days? / Pain brings on my fatigue.
Multidimensional item bank

- Knowledge about correlations between the dimensions supports selection of items having the greatest potential to reduce the statistical uncertainty about the fatigue level of the patient.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Severity</th>
<th>Impact</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity</td>
<td>1.000</td>
<td>0.495</td>
<td>0.247</td>
</tr>
<tr>
<td>Impact</td>
<td></td>
<td>1.000</td>
<td>0.580</td>
</tr>
<tr>
<td>Variability</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

- More information about fatigue (three scores instead of one)
From calibrated item pool to CAT - I

- Collaboration with other disciplines (Research Methodology, Measurement and Data Analysis, programmer development agency)
- Preparation of item bank (items, response options, parameters)
- Functional script for item selection (e.g. item selection, data base)
- Software adapted to web-technology (ROMA)
From calibrated item pool to CAT - II

- Administration and test-environment for first test-versions

- Simulations for start- and stopping rules, standard error per dimension lowest for:
  - Test length of 20 items
  - Two random start items per dimension
  - At least five items per dimension

- Test whether algorithmic codes work as they should work

- Several check ups by researchers after delivery from programmer
Demo CAT\textsubscript{FatigueRA} – severely fatigued patient

How many days did you experience fatigue…?
… not interested in anything …
... performance of routine daily activities increases fatigue ...
... medication for rheumatism cause of fatigue ...
... difficult to articulate ...
Demo CAT_{FatigueRA} – hardly fatigued patient

How long has each episode of fatigue lasted…?
... experienced normal daily activities as stressful ...
To what degree has fatigue changed...?
... less productive than I would have wished ...
... less motivated to engage in social activities ...
Usability test

- 15 patients filled in the CAT\textsubscript{FatigueRA} while thinking aloud, short consecutive interview about their experience with the new instrument

- Patients liked the new instrument, experienced it as clear, short and innovative

Citation: “… (the CAT) really goes into the fatigue and also in a very good way. So for me it was a surprising thing to do. Then you get more the idea that it is useful to fill in. You can just fill in I was tired the last weeks (…), but this says so little about fatigue.” (D12,vrouw, 60)
Usability test and flow-chart study

- Some items of the dimension “severity” can confuse patients if selected within one test due to their similarity.

- Not every item is applicable to each patient, so adapted CAT1.0 into CAT2.0 that is now used for extensive validation study.

- Data from usability test further used for flow-chart study to check:
  - Face validity (logical selection of items)
  - Measurement precision ≤ 0.32
  - Different profiles of fatigue
Examples
Flow-chart study - conclusion

- **Face validity:**
  Items were selected in a logical way

- **Measurement precision:**
  Good for severity and impact after at least 12 items but not always even small standard error (≤ 0.32) for variability

- **Profiles of fatigue:**
  Possible to see all kind of different profiles of fatigue
CAT construction – challenges

- **Possible tension between statistics and perspective of patients**
  - Dimension “variability” less stable in psychometric terms, but important aspect as reported by patients

- **Ideally flexible stopping rule – here fixed**
  - Complex to program in our multidimensional CAT\textsubscript{FatigueRA}
  - Implement in later version (also development of CAT is adaptive, not immediately one final product)

- **Sample size for IRT analyses**
  - Many participants required, re-calibration at later stage desirable
Future steps

- Promising instrument, next step validation study of CAT<sub>FatigueRA</sub>

- Measure more precise over the whole range of fatigue with manageable number of items

- Regular measurement of fatigue at outpatient clinic and have data available in consultation immediately
Thank you for your attention!

This project is financed by the Dutch Arthritis Association.

Contact: s.nikolaus@utwente.nl