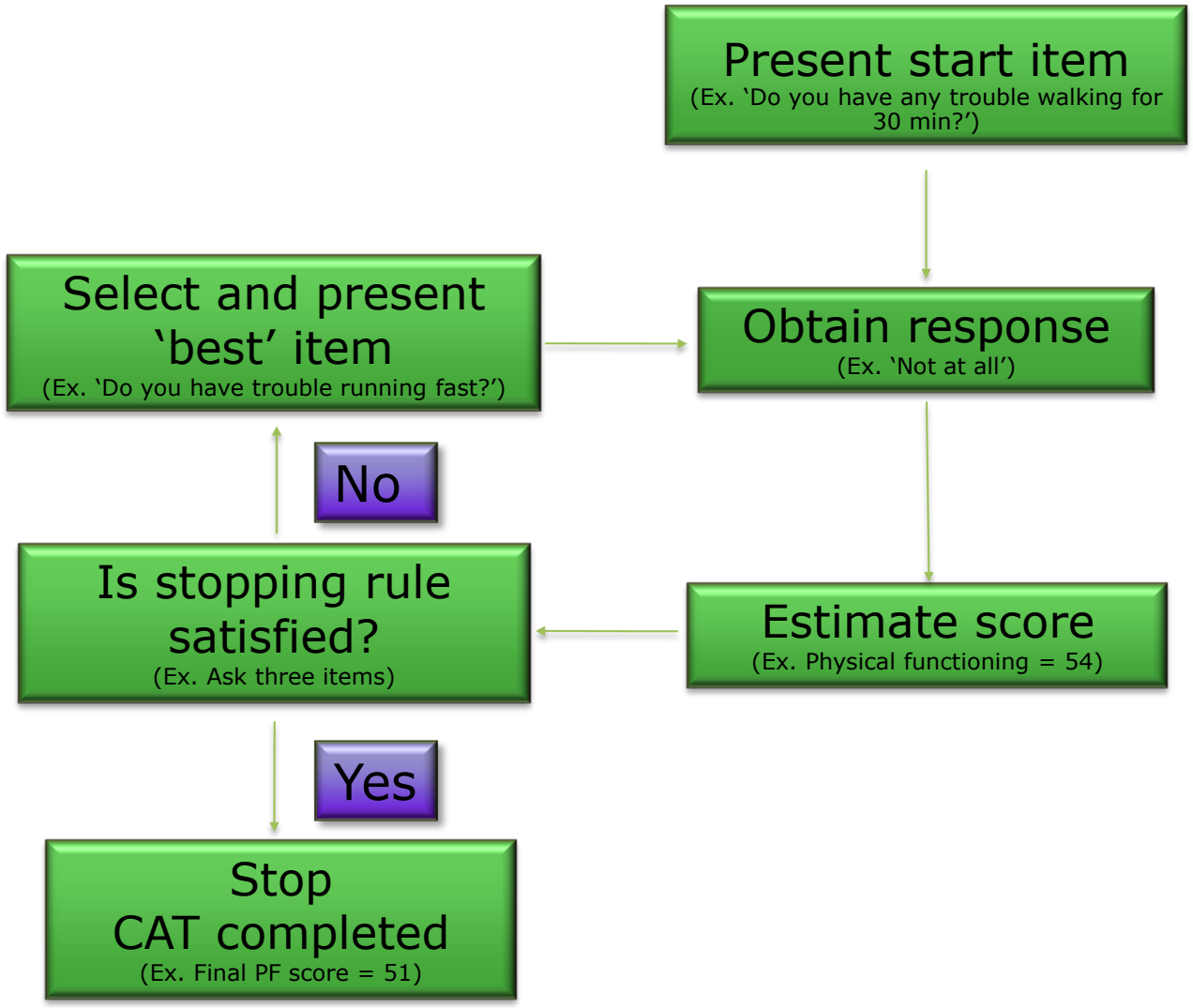
Evaluation of EUonQoL subdomains covered by dynamic CAT

Conducted by: Mogens Groenvold, Leslye Concha, & Morten Aa. Petersen

Dynamic assessment - CAT

- CAT: computerized Adaptive Testing
- CAT adapts the questionnaire, in real time, to the individual
- Aims to present the most relevant items (questions) to each individual
- Statistical model behind ensures comparability across individuals is retained
- CAT advantages include:
  - Improved measurement precision / reduced respondent burden
  - More relevant questions
  - Flexibility – can be tailored to a given aim
  - Allow for real-time scoring and feedback

CAT procedure



Background

- Eight subdomains are covered by items from EORTC CAT item banks in both the static & dynamic version of the EUonQoL-kit:
  - Physical functioning, PF
  - Role Functioning, RF
  - Fatigue, FA
  - Sleeping problems, SL
  - Emotional functioning, EF
  - Social functioning, SF
  - Pain, PA
  - Financial difficulties, FI

Background

- Eight subdomains are covered by items from EORTC CAT item banks in both the static & dynamic version of the EUonQoL-kit:
  - Physical functioning, PF (2-3)
  - Role Functioning, RF (2)
  - Fatigue, FA (2)
  - Sleeping problems, SL (1)
  - Emotional functioning, EF (3)
  - Social functioning, SF (3)
  - Pain, PA (2)
  - Financial difficulties, FI (1-2)
- **Static version**: Includes 1-3 items per subdomain, with data collected from all participants

Background

- Eight subdomains are covered by items from EORTC CAT item banks in both the static & dynamic version of the EUonQoL-kit:
  - Physical functioning, PF (7)
  - Role Functioning, RF (6)
  - Fatigue, FA (6)
  - Sleeping problems, SL (6)
  - Emotional functioning, EF (7)
  - Social functioning, SF (6)
  - Pain, PA (6)
  - Financial difficulties, FI (6)
- Static version: Includes 1-3 items per subdomain, with data collected from all participants
- **Dynamic CAT version**: Asks 6-7 items per subdomain for precise assessment ('gold standard'), collected in random 10% of sample

Aims of the evaluations

To find the optimal balance of content, precision, and response burden, i.e., to assess whether

<u>Static version</u>: we have selected the right items and the right number of items

<u>Dynamic version</u>: we have selected the right start items and when to stop (number of items/precision)

Static short forms

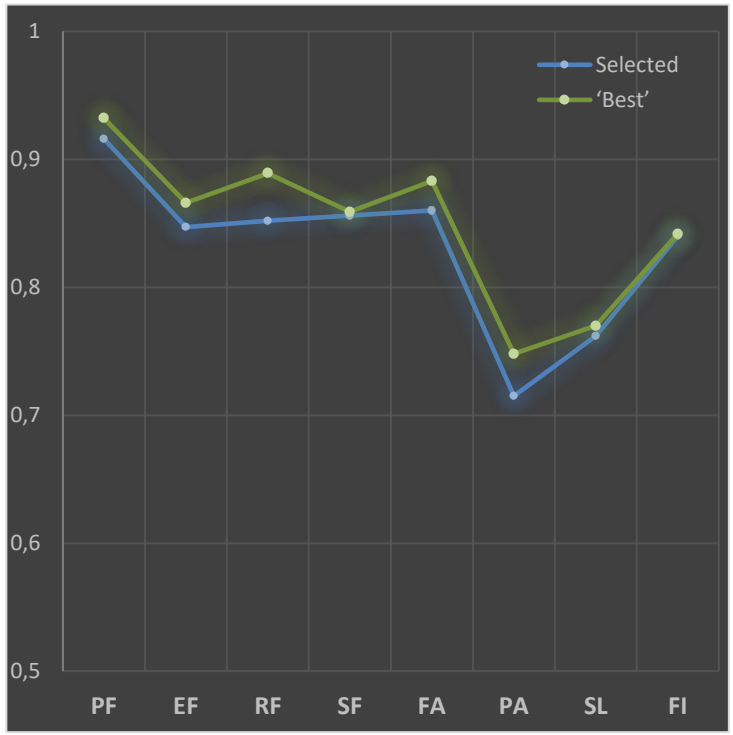
- Have the right items and the right number of items been selected?

Research questions

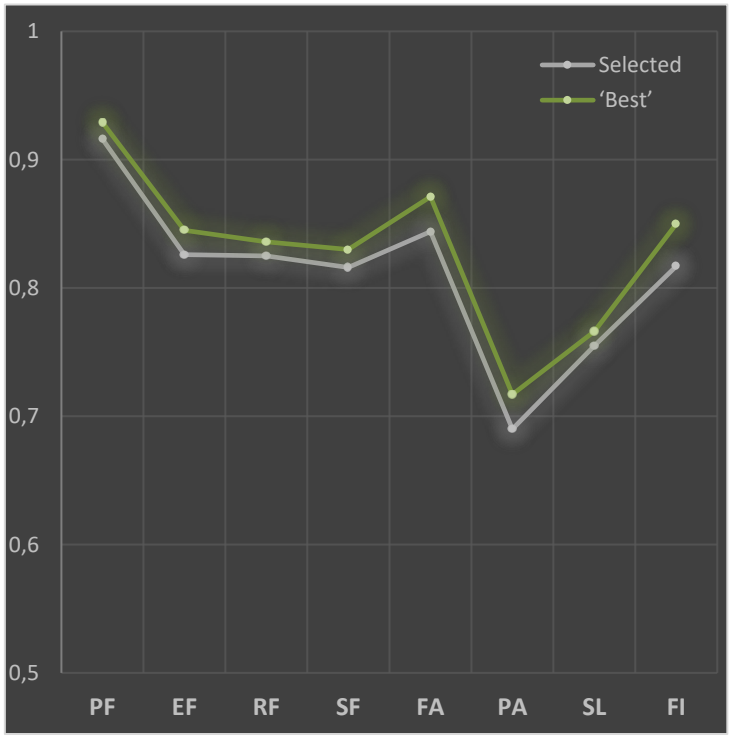
- Do the short forms provide reliable assessment?
  - - can reliability be improved?
- Is the content of the short forms relevant?
- Do the short forms provide unbiased scores?
- Do the short forms provide scores comparable to the ‘gold standard’ CATs?
- Can response burden be reduced without (relevant) reduction in measurement precision?

Short form reliability

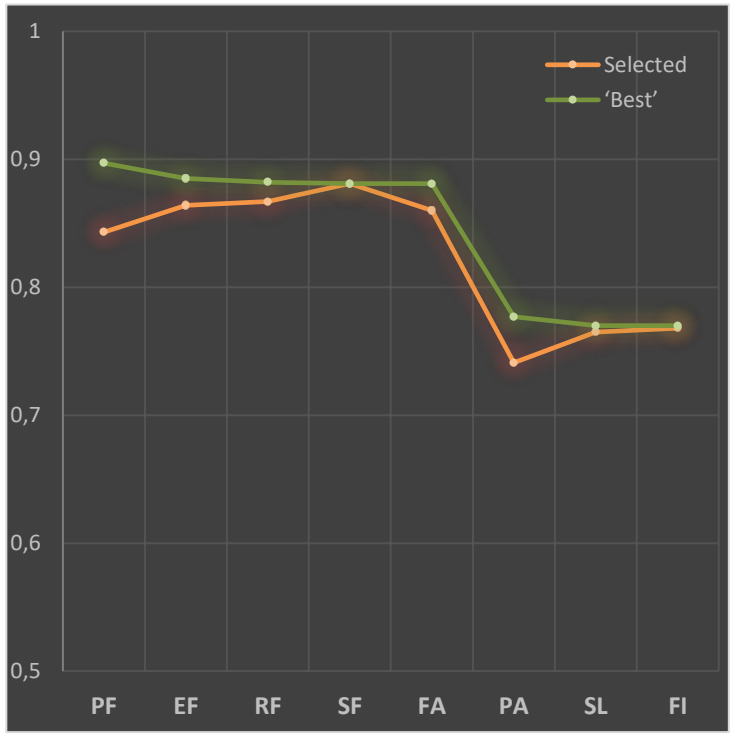
In active treatment



Survivors



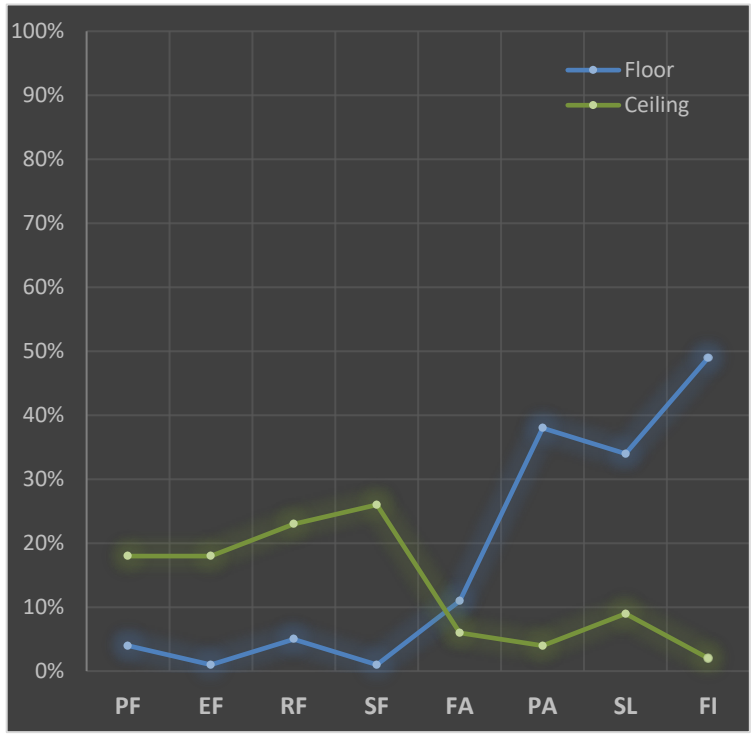
In need of palliative care



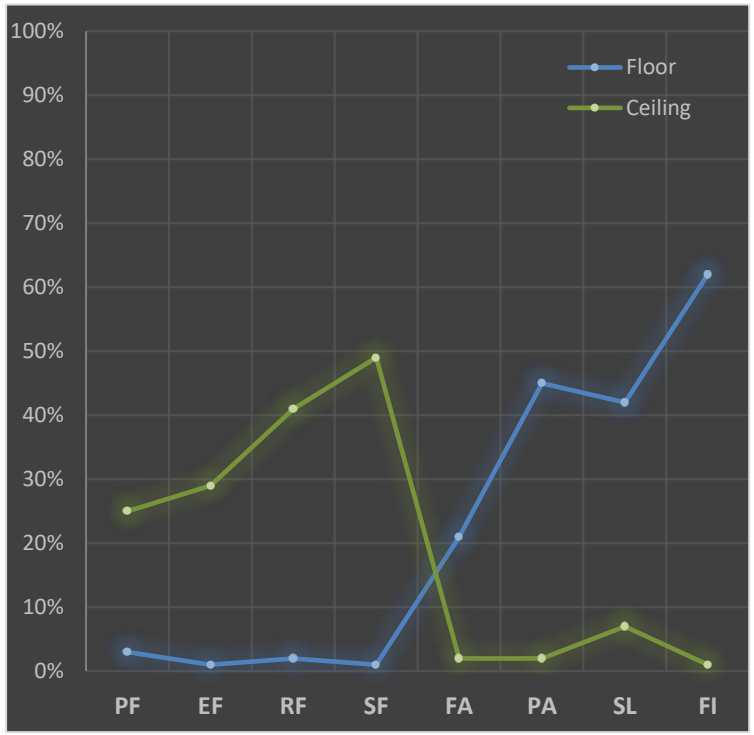
Mean reliability of selected short forms & short forms providing the highest reliability (green) for the populations

Short form floor & ceiling effects

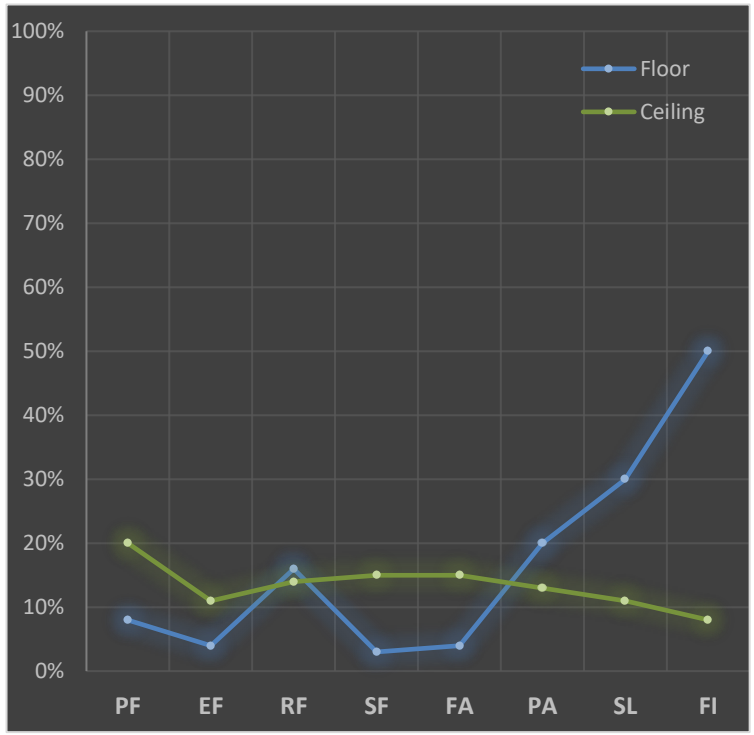
In active treatment



Survivors



In need of palliative care



Floor: Percent in lowest score; Ceiling: Percent in highest score

Short form vs. CAT – bias?

Mean differences between short forms & CATs asking 6/7 items

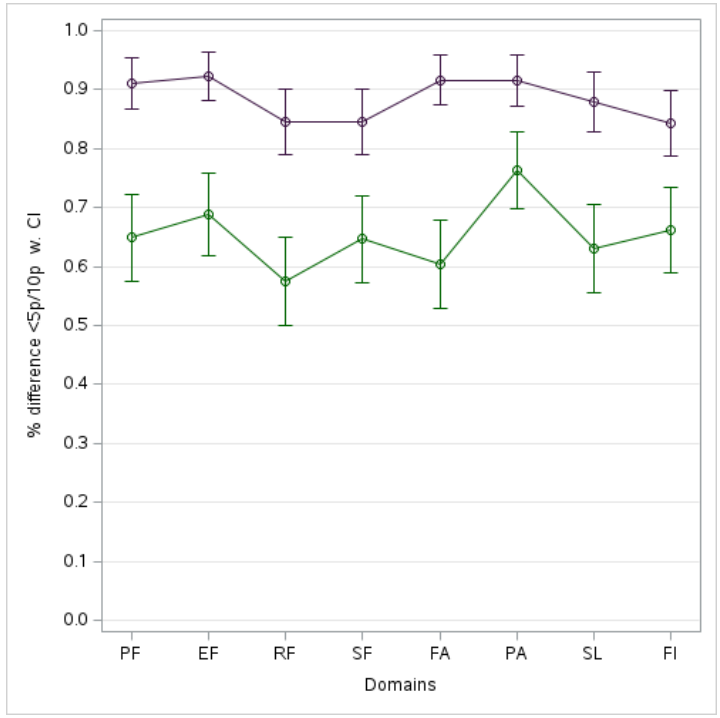
(Active= In active treatment; Palliative= In need of palliative care)

Short form vs. CAT - correlations

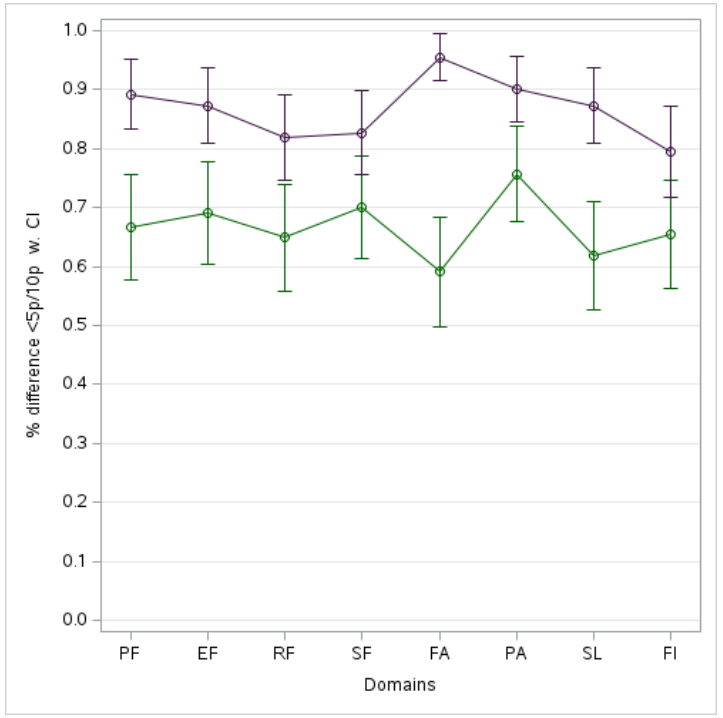
(Active= In active treatment; Palliative= In need of palliative care)

Short form vs. CAT - % diff. <5p/10p

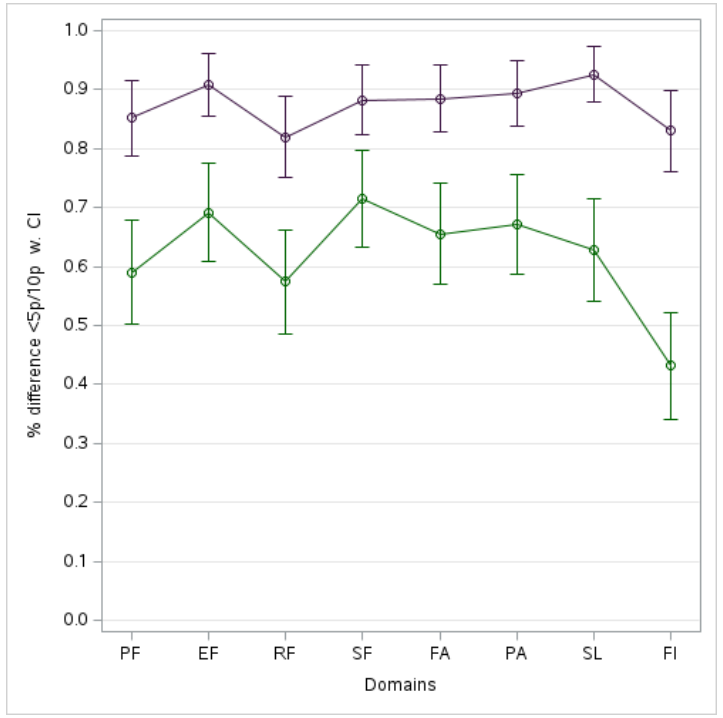
In active treatment



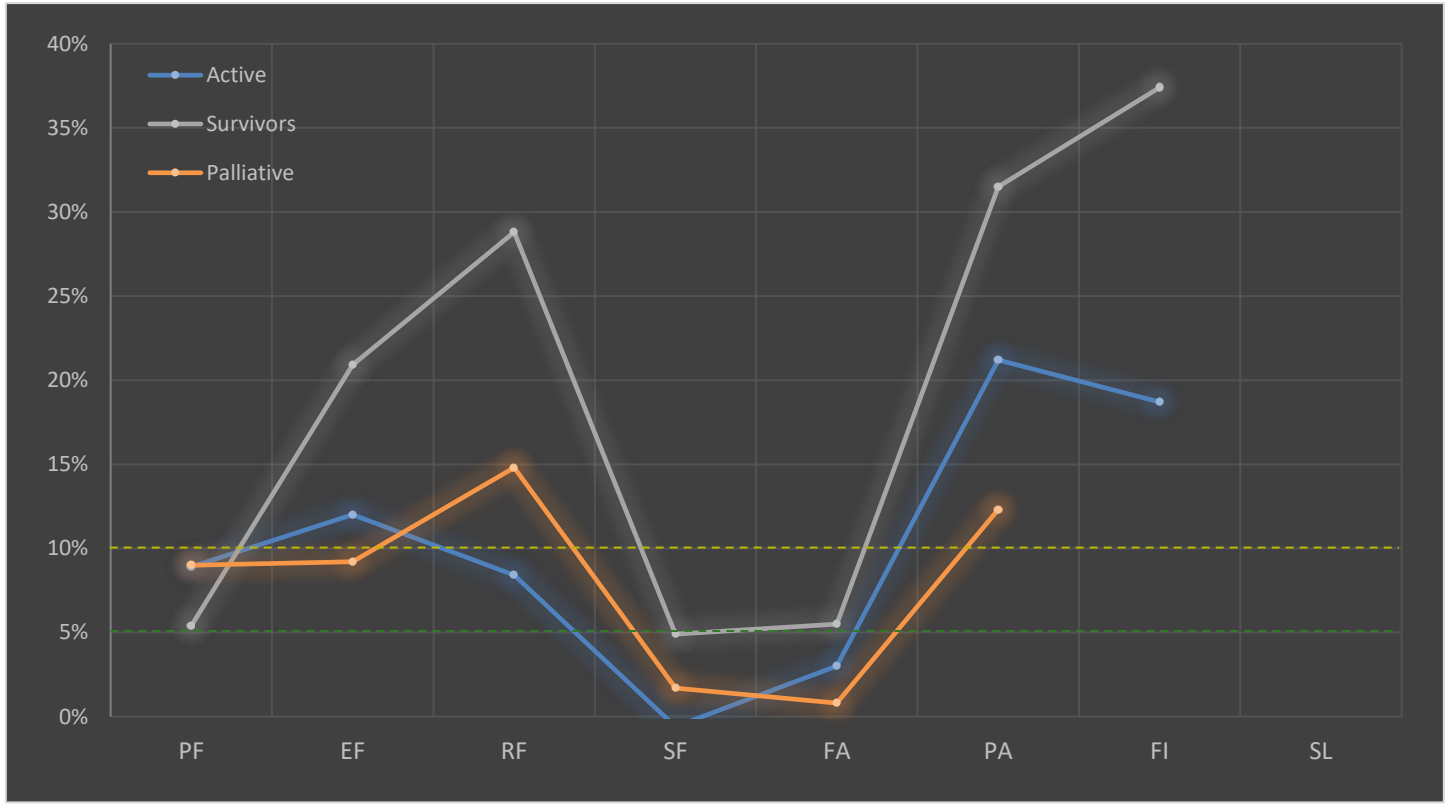
Survivors



In need of palliative care



‘Brief’ vs. current short forms



Brief short form: deleted the item providing least information from each short form
Percent additional scores deviating >5 points from CAT score when using the brief version
(Active= In active treatment; Palliative= In need of palliative care)

Short forms - conclusions

- Reliabilities of selected short forms are ‘acceptable’ and close to ‘psychometric best’ short forms
- Generally, at most trivial bias using the selected short forms
- Strong association, i.e., comparable scores, in most cases between short forms and ‘gold standard’ CAT
- Short form scores typically close to CATs – typically >60% within 5p, >80% within 10p
- Score precision reduced in most cases if abbreviating short forms – may abbreviate SF and FA (and PF?) with limited loss

Dynamic CAT

- Have the right start items been selected and when to stop?

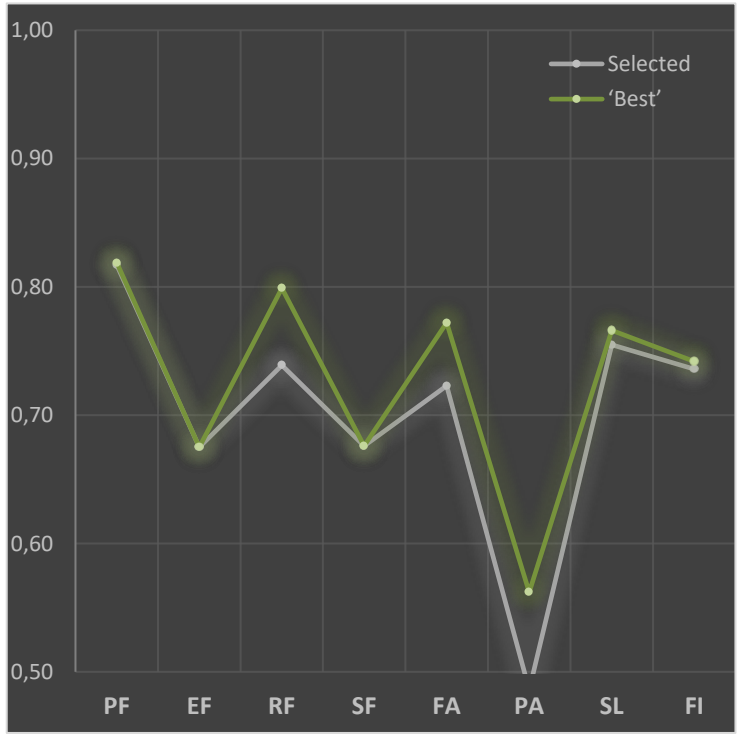
Research questions

- Do the start items provide reliable assessment?
  - - can reliability be improved?
- How reliable and precise are different length CATs?
  - – what seems the optimal length of fixed length CATs?
- How many items are needed to obtain different levels of reliability?
  - – what seems the optimal reliability for variable length CATs?

CAT - start item reliability

In active treatment

Survivors

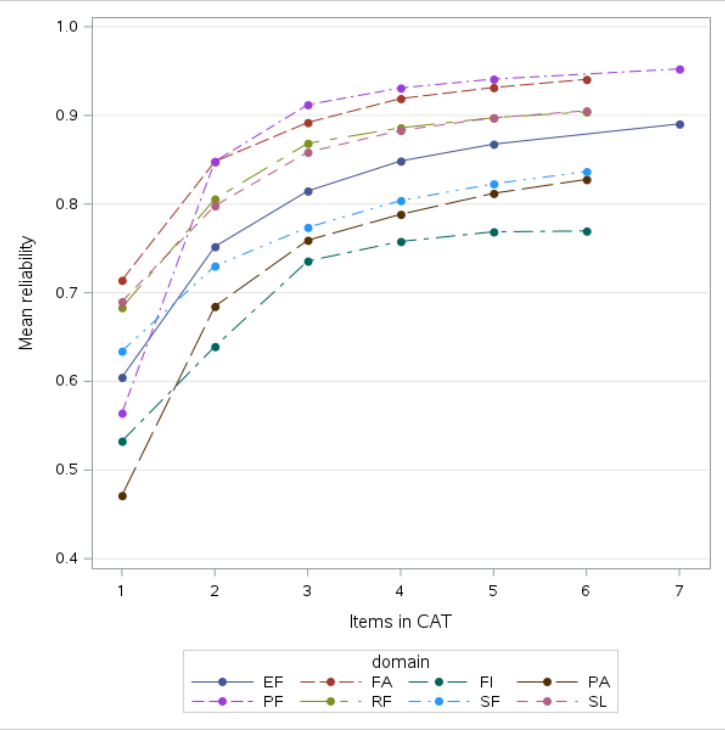


In need of palliative care

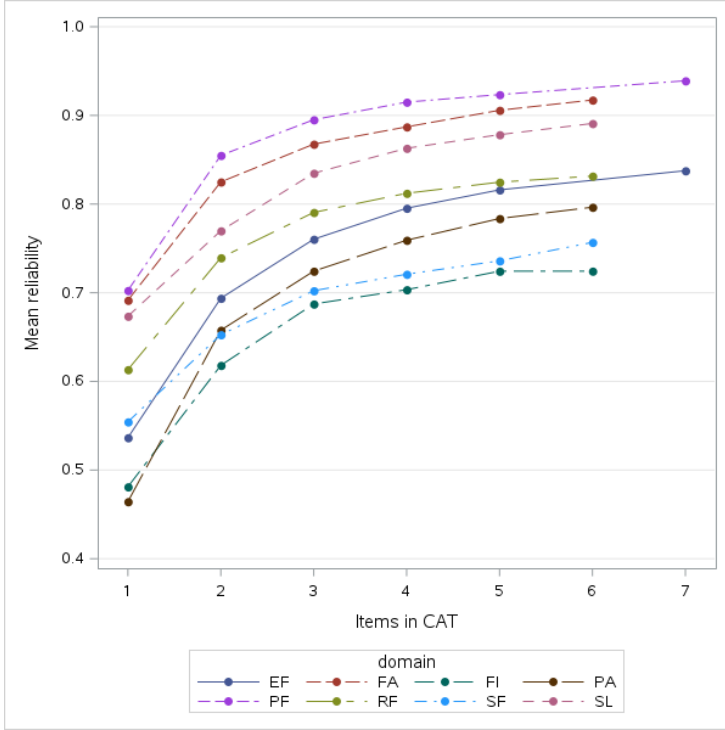
Mean reliability of the selected start items & items providing the highest reliability for the populations

Different length CATs - reliability

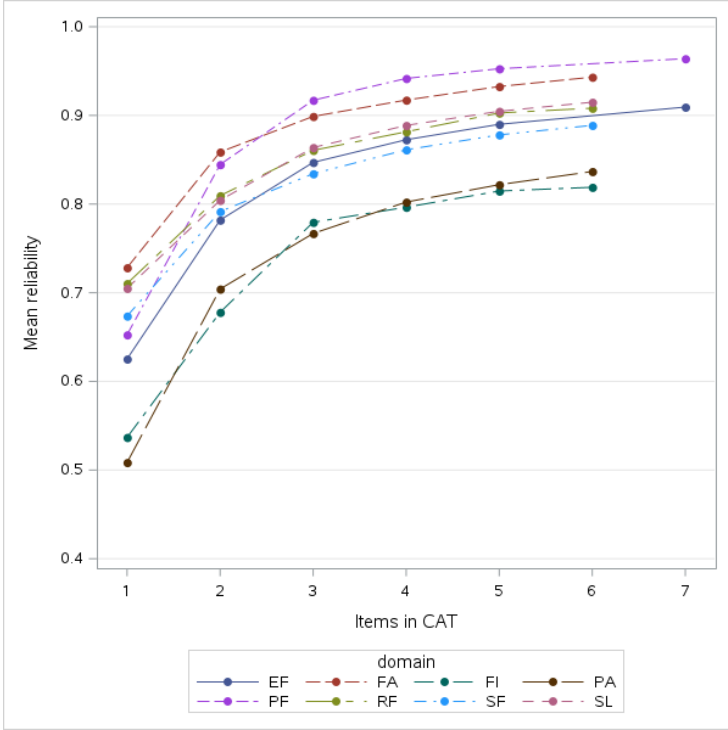
In active treatment



Survivors

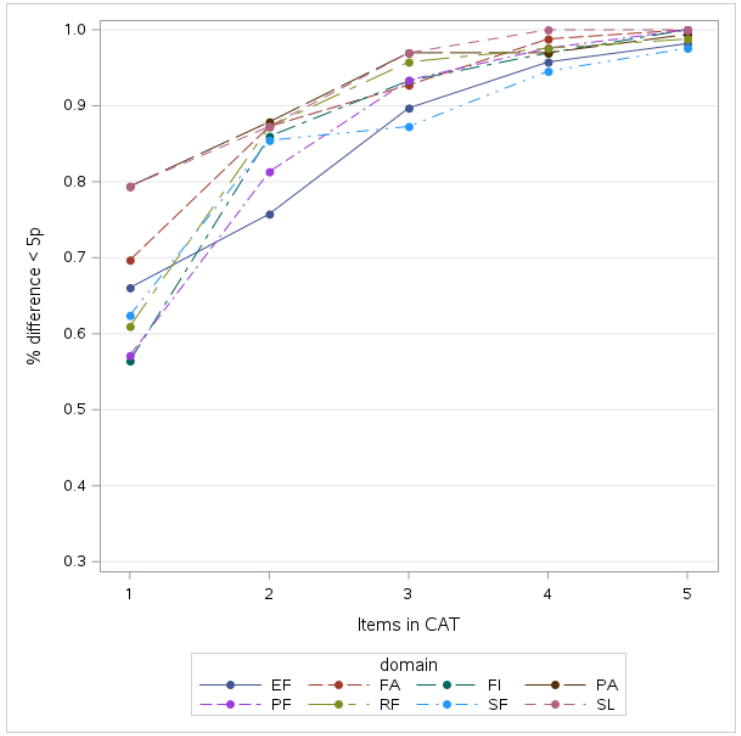


In need of palliative care

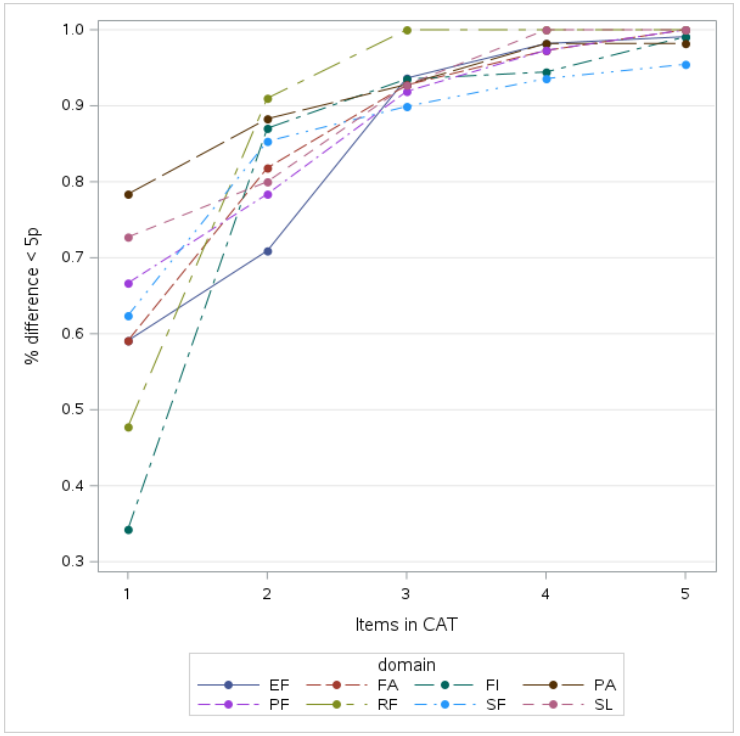


Different length CATs - % diff. <5p

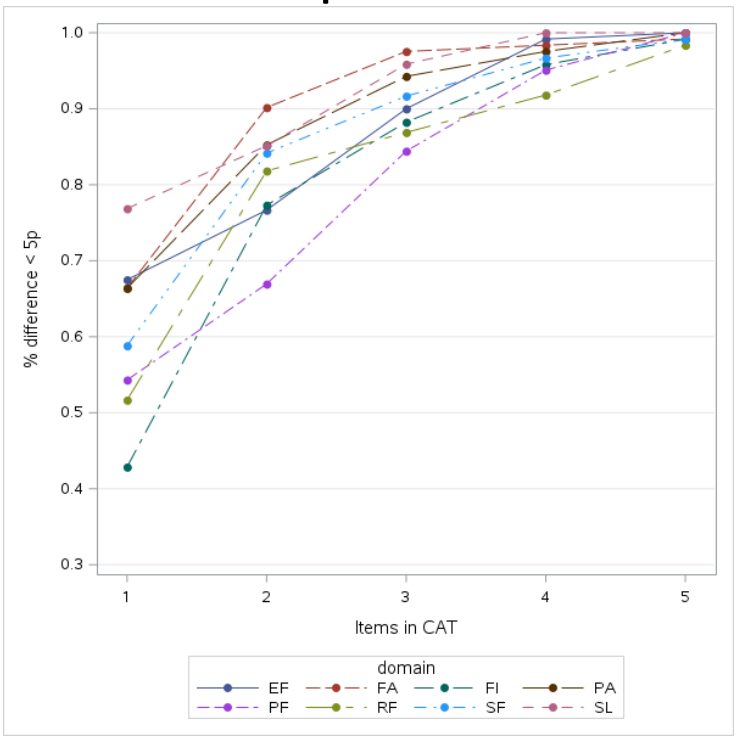
In active treatment



Survivors



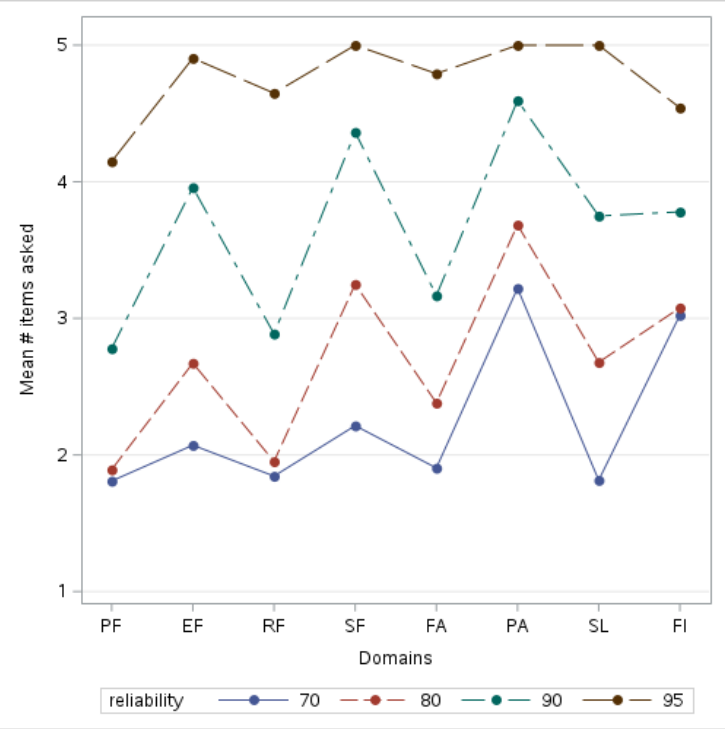
In need of palliative care



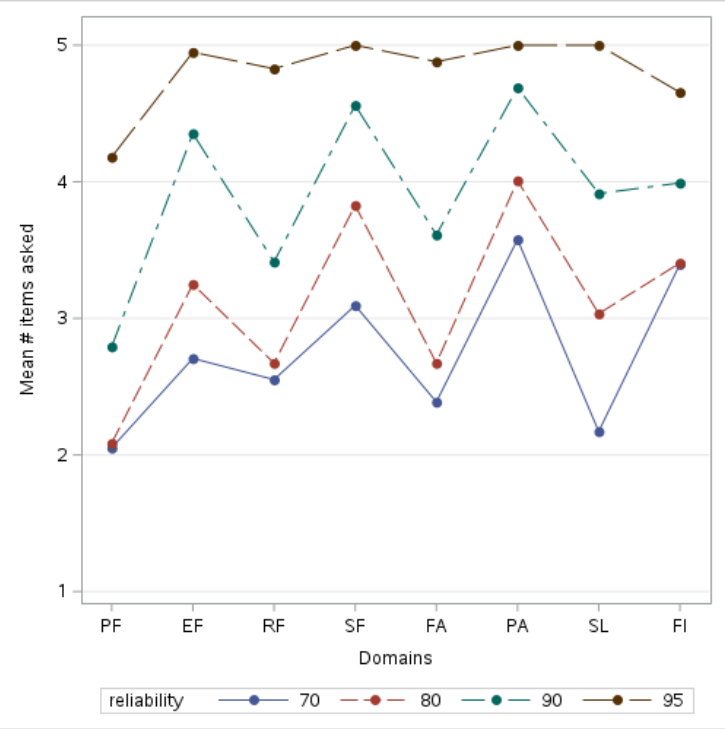
Percent scores deviating <5 points from full CAT asking all 6/7 items

Variable length CATs

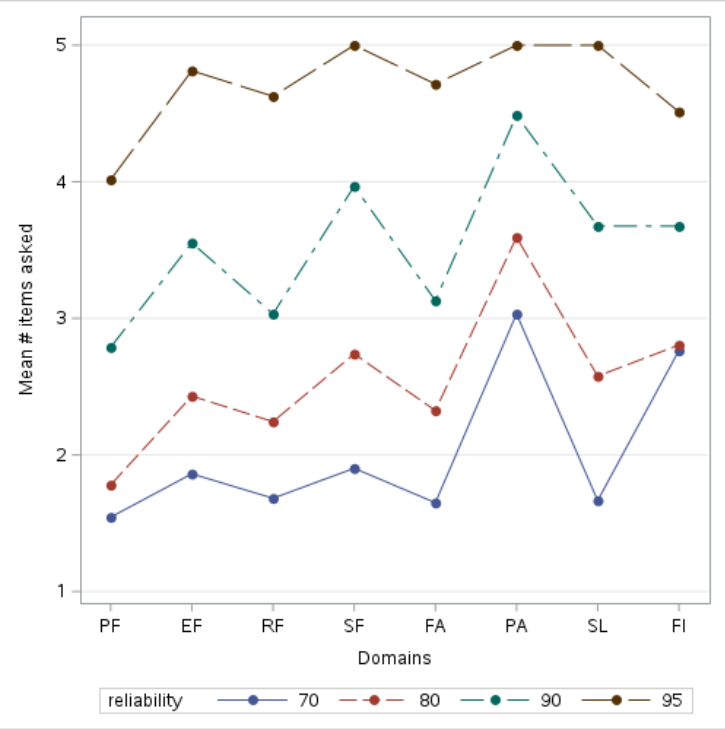
In active treatment



Survivors



In need of palliative care



Variable length CATs aiming for reliability of 0.70, 0.80, 0.90, and 0.95, respectively. Ask maximum of 5 items.

CAT - conclusions

- Reliability of selected start items typically close to ‘psychometric best’ items – some (e.g., PA) could be improved but may reduce ‘content validity’
- Fixed length CATs: asking more than three items provides limited additional precision, asking one item too imprecise -> asking two or three items may provide optimal balance between precision and efficiency – particularly if brevity is in focus
- Variable length CATs: For optimal balance between precision and efficiency these should likely aim for reliability of 0.70 - 0.80