

A Study to Assess the Inter- and Intra-rater Reliability of the 5-domain NPCCSS

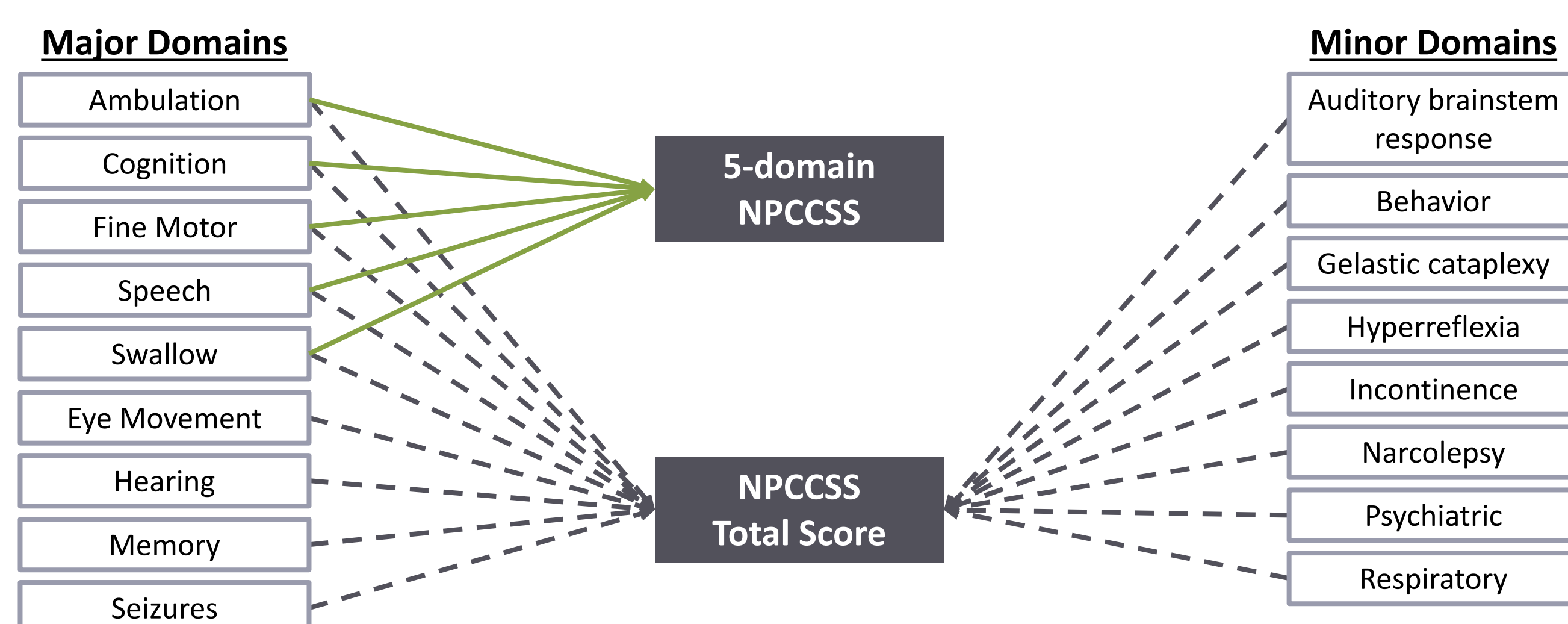
Marc Patterson¹, Christine I Dali², Tara Symonds³, Christina Guldborg², Stacie Hudgens⁴, Eugen Mengel⁵

1) Mayo Clinic, Rochester, USA; 2) Orphazyme A/S, Copenhagen, Denmark; 3) Clinical Outcomes Solutions, Folkestone, UK; 4) Clinical Outcomes Solutions, Tucson, USA; 5) Villa Metabolica Mainz; Mainz; Germany

INTRODUCTION

- NPC is a rare, chronic neurological disease caused by mutations in the NPC1 or NPC2 gene. NPC affects infants, juveniles and adults with an estimated incidence of 1:100,000 live births¹⁻⁶
- NPC is characterized by progressive and disabling neurological symptoms⁴ (Figure 1)
- The 17 domain NPC Clinical Severity Score (NPCCSS) is an established disease-specific clinician-reported outcome measure to inform disease progression in NPC patients⁷
- An abridged version, 5-domain NPCCSS, was developed comprising five domains selected by NPC patients, caregivers and experts as the most clinically relevant: ambulation, speech, cognition, fine motor skills, and swallow
 - High correlation (Spearman's correlation = 0.93) is found between the 5-domain NPCCSS and the 17-domain NPCCSS⁸
 - Establishing inter and intra-rater reliability is a necessary step in the minimization of measurement error^{7,9}

Figure 1. Conceptual Model



OBJECTIVE

- Measure inter- and intra-rater reliability of the 5-domain NPCCSS, total score and individual items as a step to verify the validation of the 5-domain NPCCSS to be used as an instrument of assessment in NPC patients

METHODS

Inclusion Criteria

- NPC1 or NPC2 patients of either gender (aged 8-, 10-, 12-, and 17-years) with at least one neurological sign of disease and ability to walk independently or with assistance were used to develop the video case studies for clinicians to rate in the inter- and intra-rater reliability exercise
- Caregiver and patient fluent in English

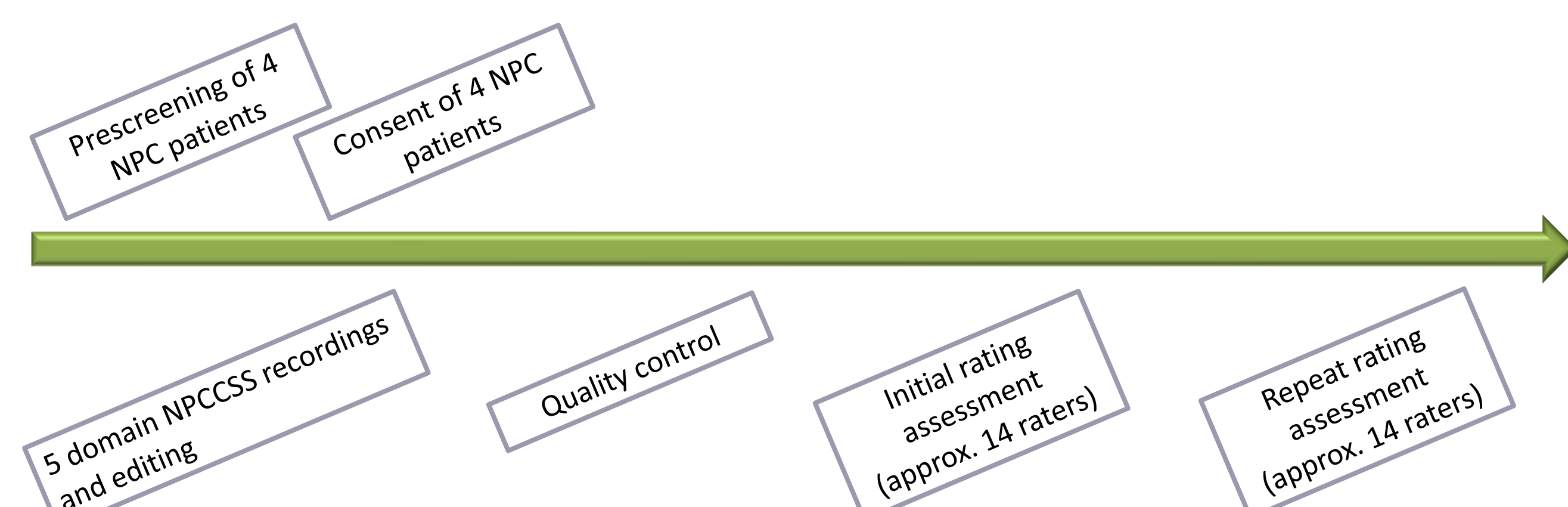
Study Design

- Study design is schematically shown in Figure 2
- Medical interview and physical examination of patients and caregivers were conducted by a medical expert in NPC; these were video-recorded and format was standardized by following a prespecified story board
- Videos were used to assess inter-rater and intra-rater reliability by a group of 13 clinicians, who were blinded to other raters' scores; raters did an initial scoring for the 5-domain NPCCSS for each patient; scoring was repeated, in a random fashion, after at least 3 weeks

Statistical Analysis

- Analytical data set incorporates initial and repeat ratings from each of the 13 raters for each of the four patient videos
- Results are reported for all time points where there was sufficient sample size to allow for meaningful interpretation of findings
- Inter- and intra-rater reliability was assessed using intra-class correlation (ICC) coefficients (score level) and kappa statistics (item level) at a 0.70 acceptability criteria threshold¹⁰
- As a sensitivity analysis to identify potential error in the reliability estimates due to rater bias, a Many-FACETS item response analysis was conducted which allows the clinician rating of severity to be derived using a standardized scale similar to the patient and the item (ie, equating to a similar metric).¹¹

Figure 2. Overview of Study Design



CGI-S = Clinician Global Impressions of Severity; NPC = Niemann-Pick disease type C; NPCCSS = Niemann-Pick type C Clinical Severity Score

RESULTS

Evaluation of 5-domain NPCCSS

- The intraclass coefficients agreement statistics associated with the evaluation of the 5-domain NPCCSS were high between the initial and repeat timepoints (intra-rater reliability $ICC_{(2,13)} = 0.937$), as well as between clinician ratings ($ICC_{(2,1)} = 0.995$)
- For the individual domains, the coefficients were above the 0.70 criterion threshold for all domains at the initial timepoint (range 0.763–0.954)
- At the repeat timepoint, ambulation fell slightly below the criterion threshold (coefficient of concordance = 0.681), with all other domains above the criterion threshold (range 0.723–0.949) (Table 1)
- As a sensitivity analysis, an evaluation of the clinician ratings, patient videos, and 5-domain NPCCSS using the item-levels FACETS analysis, indicated strong agreement across the clinician group (57.3% exact agreement; 0.87 correlation across ratings; sample error = 0.01)
 - In terms of patient videos utility, there were no videos which performed poorly (eg, all videos fit the model, none were misfitting; infit value ≥ 1.5 ; range 0.76–1.11) and the correlation across ratings was 0.73 with an error rate of 0.02
- For the individual domains, there were no misfitting domains (infit value ≥ 1.5 ; range 0.72–1.16) and the correlation across ratings was 0.85 with an error rate of 0.01
- Rating scale analysis, conducted using FACETS, an extension of Rasch measurement for judge-mediated data, indicated monotonic increases in severity rating from lowest (verbal response option = 0; average person measure = -4.59) to most severe (verbal response option = 5; average person measure = 3.53)
- Agreement between the patient video severity ratings was observed between the classical descriptive statistics on the 5-domain NPCCSS and the item-level FACETS analysis. In each analysis, Subject C was the most severe (NPCCSS mean: initial 18.15, repeat 18.85; FACETS 3.30) and Subject D was the least severe (NPCCSS mean: initial 4.46, repeat 4.77; FACETS -2.20).

Table 1. Item Level Coefficient of Concordance

Visit	Variable	Kappa Coefficient of Concordance
Initial	Ambulation Score	0.785
	Speech Score	0.954
	Swallow Score	0.813
	Fine Motor Skills Score	0.763
	Cognition Score	0.917
Repeat	Ambulation Score	0.681
	Speech Score	0.949
	Swallow Score	0.858
	Fine Motor Skills Score	0.723
	Cognition Score	0.883

CONCLUSIONS

- The inter- and intra-rater reliability analyses suggest that clinicians were able to rate patients similarly and consistently, as well as designate each of the categorical severity ratings in the manner intended:
 - More specifically, clinicians were able to utilize the measure in the manner intended, whilst reviewing multiple videos representing each level of clinical severity
 - Patients with less severity were rated with lower categories on the 5-domain NPCCSS by participating clinicians
- The supportive FACETS Rasch analysis illustrated that each of the items contributing to the 5-domain NPCCSS was locally independent and that the instrument is able to target patients along the severity continuum with complementary precision
- Based on this analysis, the 5-domain NPCCSS measurement tool should be considered both valid and reliable as an endpoint in future studies

REFERENCES

1. Huang JY, et al. *J Formosan Med Assoc.* 2011;110(8):537-542.
2. Imrie J, et al. *BMC Neurol.* 2015;15(15):257.
3. Patterson MC, et al. *Neurol Clin Pract.* 2017;7(6):449-511.
4. Vanier MT. *Orphanet J Rare Dis.* 2010;5(1):16.
5. Wassif CA, et al. *Genet Med.* 2016;18(1):41-48.
6. Geberhiwot T, et al. *Orphanet J Rare Dis.* 2018;13(1):50
7. Fleiss JL. *Statistical methods for rates and proportions.* 2nd ed. New York: John Wiley; 1981.
8. Cortina-Borja M, et al. *Orphanet J Rare Dis.* 2018;13(1):143.
9. Sim J, Wright C. *Research in health care: concepts, designs and methods.* Nelson Thornes; 2000.
10. Landis JR, Koch GG. *Biometrics.* 1977;33(1):159-74.
11. Linacre, J. *Many-facet Rasch measurement.* 1994, Chicago: MESA Press.

DISCLOSURES

MP: MP has received research funding and consultant honoraria from: Actelion, Agios, Amicus, IntraBio, Novartis, Orphazyme, Shire, Vtesse, NIH, Peggy Furth Fund. EM: EM current address: SphinCS GmbH, Hochheim, Germany. EM has received research funding and consultant honoraria from Actelion, Alexion, Orphazyme A/S, Sanofi Genzyme and Takeda. CGU and CID are employees of Orphazyme A/S. TS and SH are employees of Clinical Outcomes Solutions with no conflicts of interest.

ACKNOWLEDGMENTS

We are thankful to the raters, NPUK, and the patients participating on the videos. Medical writing and poster preparation assistance was provided by Clinical Outcomes Solutions (Chicago, USA). This study was funded by Orphazyme.