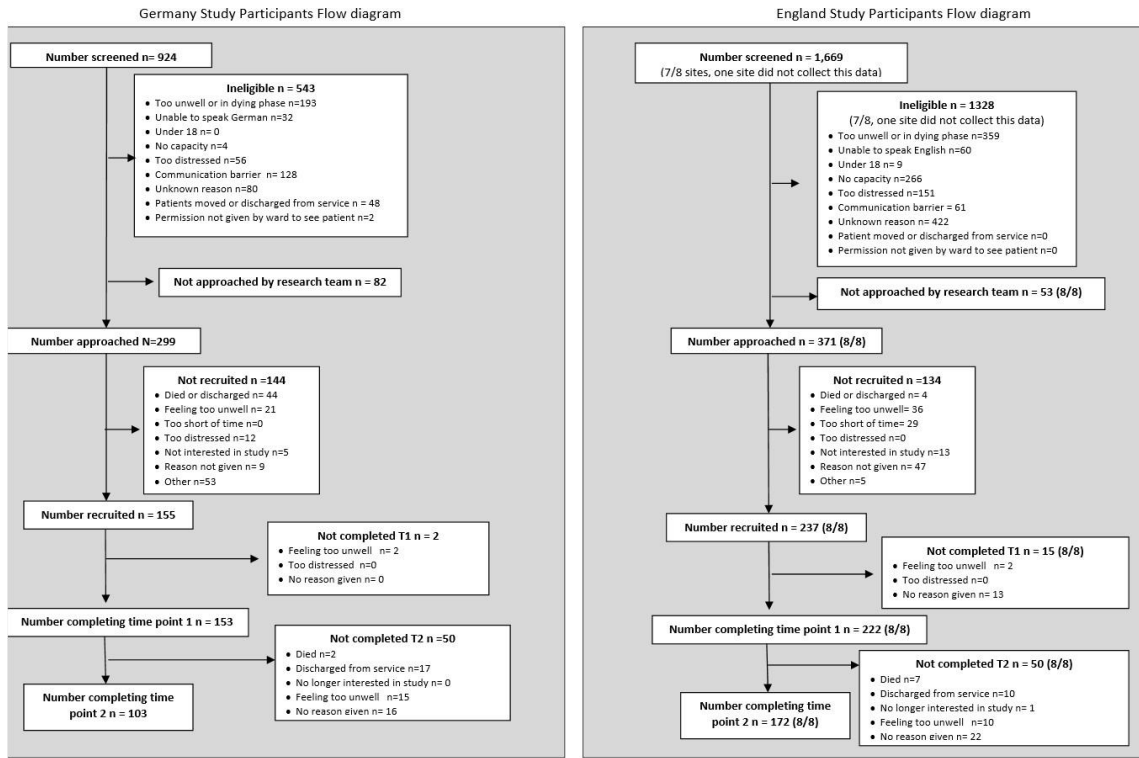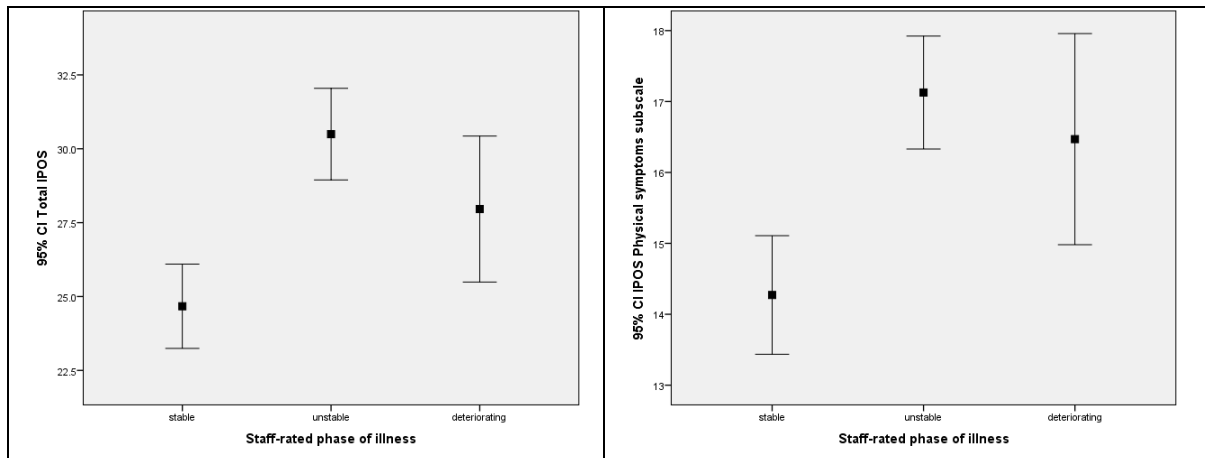**Additional Online Appendix for**

Title: A Brief, Patient- and Proxy-Reported Outcome Measure in Advanced Illness: Validity, Reliability and Responsiveness of the Integrated Patient Outcome Scale (IPOS)

Murtagh FEM, Ramsenthaler C, Firth A, Groeneveld EI, Lovell N, Simon ST, Denzel J, Bernhardt F, Schildmann EM, Oorschot B, Hodiamont F, Streitwieser S, Higginson IJ, Bausewein C

Appendix Figure 1: Flowchart of numbers of patient participants screened, eligible, approached and consented, plus numbers completing each timepoint, with reasons for non-completion

Germany Study Participants Flow diagram

Number screened n= 924

Ineligible n = 543
- Too unwell or in dying phase n=193
- Unable to speak German n=32
- Under 18 n= 0
- No capacity n=4
- Too distressed n=56
- Communication barrier n= 128
- Unknown reason n=80
- Patients moved or discharged from service n = 48
- Permission not given by ward to see patient n=2

Not approached by research team n = 82

Number approached N=299

Not recruited n =144
- Died or discharged n= 44
- Feeling too unwell n= 21
- Too short of time n=0
- Too distressed n=12
- Not interested in study n=5
- Reason not given n= 9
- Other n=53

Number recruited n = 155

Not completed T1 n = 2
- Feeling too unwell n= 2
- Too distressed n=0
- No reason given n= 0

Number completing time point 1 n = 153

Not completed T2 n =50
- Died n=2
- Discharged from service n=17
- No longer interested in study n= 0
- Feeling too unwell n=15
- No reason given n= 16

Number completing time point 2 n = 103

England Study Participants Flow diagram

Number screened n = 1,669
(7/8 sites, one site did not collect this data)

Ineligible n = 1328
(7/8, one site did not collect this data)
- Too unwell or in dying phase n=359
- Unable to speak English n=60
- Under 18 n= 9
- No capacity n=266
- Too distressed n=151
- Communication barrier = 61
- Unknown reason = 422
- Patient moved or discharged from service n=0
- Permission not given by ward to see patient n=0

Not approached by research team n = 53 (8/8)

Number approached n = 371 (8/8)

Not recruited n =134
- Died or discharged n= 4
- Feeling too unwell= 36
- Too short of time= 29
- Too distressed n=0
- Not interested in study n=13
- Reason not given n= 47
- Other n=5

Number recruited n = 237 (8/8)

Not completed T1 n = 15 (8/8)
- Feeling too unwell n= 2
- Too distressed n=0
- No reason given n= 13

Number completing time point 1 n = 222 (8/8)

Not completed T2 n = 50 (8/8)
- Died n=7
- Discharged from service n=10
- No longer interested in study n= 1
- Feeling too unwell n=10
- No reason given n= 22

Number completing time point 2 n = 172 (8/8)

Appendix Figure 2: Known group comparisons showing patient-rated total IPOS mean score and IPOS symptoms subscale total mean scores by staff-rated phase of illness (with 95% CIs).

Appendix Table 1: Results of confirmatory factor analyses

Index of fit	One-factor CFA (n=376)	Two-factor CFA (n = 376)	Three-factor CFA (n = 376)
No. of items	17	17	17
Chi-Square[a]	578.349	343.023	224.762
df	119	89	62
p-value	0.000	0.000	0.000
Chi-square/df[b]	4.9	3.9	2.5
CFI[c]	0.546	0.822	0.949
TLI[c]	0.481	0.791	0.935
RMSEA[d]	0.110 (0.101; 0.118)	0.062 (0.055; 0.069)	0.060 (0.052; 0.068)

[a]Chi-square test should be non-significant. This is usually not achieved in large samples.
[b]Ratio of Chi-square/degrees of freedom of less than 2.00 is deemed to reflect good fit of the model.
[c]Confirmatory fit index (CFI) and TLI (Tucker-Lewis-Index) of more than 0.90 is deemed as acceptable and of more than 0.95 as good fit to the model.
[d]Root mean square error of approximation (RMSEA) of less than 0.08 is deemed good fit.

Appendix Table 2: Factor loadings for all IPOS items

Item	Factor 1: Physical symptoms	Factor 2: Emotional symptoms	Factor 3: Communication/ Practical Issues
Pain	.29		
Shortness of breath	.37		
Weakness or lack of energy	.74		
Gastrointestinal parcel (Nausea, Vomiting, Constipation)	.33		
Poor appetite	.42		
Sore or dry mouth	.40		
Drowsiness	.57		
Poor mobility	.57		
Anxiety		.83	
Family anxiety		.51	
Depression		.69	
Feeling at peace		.41	
Sharing feelings			.65
Information			.73
Practical matters			.39
Eigenvalue	3.5	1.7	1.2
% variance explained	24.9	12.3	8.3

Appendix: Convergent and discriminant validity

The following correlations were hypothesised:
1. The total ESAS score will yield high correlations to both IPOS total and IPOS Physical symptoms subscale scores. Due to IPOS being a more comprehensive tool than the ESAS, a higher correlation of ESAS and IPOS Physical Symptoms than between the total ESAS score and IPOS Total Score was hypothesised.
2. High correlations between single items from the ESAS and the IPOS measuring the same symptoms were hypothesised. These symptom pairs were: pain, shortness of breath, fatigue-weakness or lack of energy, nausea, appetite loss- poor appetite, anxiety/worry, feeling depressed/depression.
3. All subscales from the FACT-G were hypothesised to have mid-range correlations with the total IPOS score.
4. The IPOS Physical symptoms subscale will yield higher correlations with physical and functional wellbeing from the FACT-G than with FACT-G emotional and social wellbeing subscales.
5. The IPOS Emotional symptoms subscale will yield higher correlations with FACT-G emotional and social well-being subscales than with functional and physical subscales.
6. The IPOS Communication/Practical Issues subscale would yield a middle-range correlation with the FACT-G social wellbeing subscale.
7. The poor mobility item on the IPOS would yield higher correlations with the FACT-G functional wellbeing subscale than with emotional and social well-being.
8. The global quality of life item on the FACT-G and the global quality of life item on the EORTC would yield mid-range correlations with the total IPOS score.

Hypothesis 1 and 2:
The total ESAS score showed strong correlations to the total IPOS and IPOS Symptoms subscale score. All single symptom items from the ESAS correlated highly with the respective symptom items on the

IPOS (see Appendix Table 3). Correlations were in the range of 0.48 (for the item 'Anxiety') to 0.74 ('Shortness of breath').

Hypothesis 3 and 4:
The subscales of the FACT-G and its total score showed small to high correlations with the total IPOS score (range between 0.08 and 0.70) (see Appendix Table 4). The lowest correlation was observed between the social subscale of the FACT-G and the total IPOS score ($r = 0.08$), as hypothesised due to the lack of overlap between the two scales. The FACT-G physical subscale showed high correlations with the IPOS Physical symptoms subscale ($r = 0.64$), similarly but not as high correlation with the IPOS subscale Emotional symptoms ($r = 0.58$).

Appendix Table 3: Correlations between IPOS single symptom items and the corresponding Edmonton Symptom Assessment Tool items (n = 347)

	r	**95% CI**
Pain	0.59	0.51 – 0.67
Weakness or lack of energy/Tiredness	0.52	0.44 – 0.60
Drowsiness	0.50	0.43 – 0.57
Nausea	0.63	0.51 – 0.75
Poor appetite/Lack of appetite	0.68	0.62 – 0.74
Shortness of breath	0.74	0.69 – 0.79
Depression	0.61	0.54 – 0.68
Anxiety or worry about illness or treatment/Anxiety	0.48	0.40 – 0.56

Hypothesis 5 and 6:
The same pattern was found for the emotional and social subscales of the FACT-G that correlated more strongly with the Emotional symptoms subscale of the IPOS than with the Physical symptoms subscale. The IPOS Communication/Practical Issues subscale showed only a small correlation with the FACT-G social wellbeing subscale, indicating that the IPOS Communication/Practical Issues subscale measures unique aspects not represented in any other measure, as was apparent in the general small to non-existing correlations of this subscale with the ESAS, FACT-G and EORTC subscales.

Hypothesis 7 and 8:
The IPOS poor mobility item likewise correlated more strongly with the physical and functional subscale of the FACT-G than with its social and emotional subscales. Both the FACT-G and the EORTC QLQ-C30 global quality of life items showed mid-range correlations to the total IPOS score ($r = -0.47$ and -0.30, respectively), highlighting that the IPOS measures more aspects than either of these global items.

**Appendix Table 4: Correlations between IPOS, ESAS and FACT-G at baseline
(hypothesised associations in bold)**

Item/ Subscale	n	Total IPOS	IPOS Physical symptoms	IPOS Emotional symptoms	IPOS Communication /Practical Issues	IPOS Poor mobility
FACT-G Total	341	**.68**	**.45**	**.62**	**.32**	**.35**
Physical	361	**.70**	**.64**	.42	.48	**.45**
Social	352	.08	.05	**.03**	**.20**	.10
Emotional	357	**.56**	**.33**	**.58**	.18	.26
Functional	355	**.51**	**.36**	.45	.23	**.34**
FACT-G QoL	350	**-.47**	.33	.41	.20	-.30
ESAS Total	346	**.68**	**.57**	.47	.23	.38
EORTC						
Global QOL	352	**-.30**	-.21	-.34	-.10	-.32

IPOS: Integrated Patient Outcome Scale, ESAS: Edmonton Symptom Assessment Tool, FACT-G: Functional Assessment of Cancer Therapy-G, EORTC: European Organization for Research and Therapy of Cancer Quality of Life Questionnaire C30.

**Appendix Table 5: Known-group comparisons for IPOS total and IPOS subscale scores,
comparing participants with low and high Australia-modified Karnofsky performance status and participants with different Phases of illness**

	Australia-modified Karnofsky performance status				Phase of illness				
	60-100 (n=126)	0-50 (n=75)	t	p-value	Stable (n = 164)	Unstable (n = 129)	Deteriorating (n = 52)	F	p
IPOS Total	26.2 ± 9.6	29.0 ± 8.8	2.760	**.006**	24.7 ± 9.2	30.5 ± 8.8	28.0 ± 8.9	15.09	**.000**
IPOS Physical	14.8 ± 6.1	17.2 ± 5.7	3.775	**.000**	13.9 ± 5.9	17.8 ± 5.4	16.9 ± 6.4	17.76	**.000**
IPOS Emotional	7.8 ± 3.5	8.7 ± 3.6	2.346	**.020**	7.5 ± 3.7	8.9 ± 3.5	8.4 ± 3.2	5.76	**.030**
IPOS Communication /Practical Issues	3.6 ± 2.7	3.1 ± 2.7	-1.915	.056	3.3 ± 2.7	3.7 ± 2.9	2.9 ± 2.2	1.89	.153
Items									
Pain	1.8 ± 1.2	1.9 ± 1.2	.231	.818	1.6 ± 1.1	2.0 ± 1.2	1.8 ± 1.2	4.50	**.012**
Shortness of breath	1.2 ± 1.1	1.5 ± 1.3	2.472	**.014**	1.1 ± 1.1	1.5 ± 1.2	1.6 ± 1.3	5.82	**.003**
Weakness or lack of energy	2.3 ± 1.0	2.6 ± 1.0	2.931	**.004**	2.2 ± 1.0	2.6 ± 1.0	2.6 ± 0.9	7.09	**.001**
Nausea	1.0 ± 1.1	1.1 ± 1.2	.855	.393	0.8 ± 1.1	1.2 ± 1.2	1.0 ± 1.2	6.19	**.002**
Vomiting	0.4 ± 0.9	0.6 ± 1.1	1.267	.206	0.4 ± 0.9	0.7 ± 1.1	0.4 ± 0.9	4.17	**.016**
Poor appetite	1.5 ± 1.3	1.7 ± 1.3	1.594	.112	1.3 ± 1.2	1.9 ± 1.2	1.8 ± 1.3	9.21	**.000**
Constipation	1.3 ± 1.3	1.3 ± 1.3	.074	.941	1.2 ± 1.3	1.4 ± 1.3	1.4 ± 1.3	1.33	.266
Sore or dry mouth	1.7 ± 1.2	1.8 ± 1.3	.757	.450	1.7 ± 1.2	2.1 ± 1.1	2.1 ± 1.0	5.20	**.006**
Drowsiness	1.8 ± 1.1	2.1 ± 1.1	2.490	**.013**	1.6 ± 1.3	2.0 ± 1.2	1.7 ± 1.3	3.96	**.020**
Poor mobility	2.1 ± 1.2	2.9 ± 1.1	6.611	**.000**	2.3 ± 1.3	2.6 ± 1.1	2.7 ± 1.2	2.59	.077
Patient anxiety	2.1 ± 1.3	2.3 ± 1.3	1.888	.060	1.9 ± 1.3	2.4 ± 1.2	2.2 ± 1.2	4.60	**.011**
Family anxiety	2.7 ± 1.2	3.1 ± 1.1	2.810	**.005**	2.7 ± 1.3	3.1 ± 1.0	2.9 ± 0.9	4.90	**.008**
Depression	1.5 ± 1.3	1.7 ± 1.3	2.088	**.037**	1.4 ± 1.3	1.8 ± 1.3	1.6 ± 1.2	2.65	.072
Feeling at peace	1.6 ± 1.2	1.6 ± 1.2	.177	.860	1.5 ± 1.2	1.8 ± 1.2	1.7 ± 1.2	2.09	.125
Sharing feelings	1.4 ± 1.3	1.2 ± 1.3	-1.831	.068	1.3 ± 1.3	1.4 ± 1.4	1.1 ± 1.2	1.00	.368
Information	1.2 ± 1.2	0.9 ± 1.2	-2.127	**.034**	1.1 ± 1.2	1.2 ± 1.2	0.8 ± 1.0	1.65	.194
Practical matters	1.1 ± 1.2	1.0 ± 1.2	-.192	.848	1.0 ± 1.2	1.2 ± 1.3	1.0 ± 1.1	1.03	.359

Appendix Table 6: Inter-rater agreement between two independent staff members. ICC$_{agreement}$ with 95% CI and standard error of measurement SEM$_{agreement}$ for all IPOS items using variance component method

	IPOS Item	Var (patient)	Var (rater)	Var (error)	ICC$_a$	Lower CI	Upper CI	SEM$_a$
1	Pain	1.28	0.00	1.49	0.72	0.61	0.81	1.22
2	Shortness of breath	1.49	0.00	1.66	0.80	0.71	0.86	1.29
3	Weakness or lack of energy	0.21	0.00	0.61	0.25	0.05	0.44	0.78
4	Nausea	0.61	0.00	1.25	0.63	0.49	0.74	1.12
5	Vomiting	0.79	0.00	0.97	0.61	0.47	0.73	0.99
6	Poor appetite	0.92	0.01	1.18	0.46	0.29	0.61	1.09
7	Constipation	0.96	0.01	1.32	0.41	0.21	0.57	1.15
8	Sore or dry mouth	0.89	0.00	1.14	0.49	0.31	0.64	1.07
9	Drowsiness	0.20	0.00	0.76	0.21	0.01	0.41	0.87
10	Poor mobility	0.56	0.00	0.66	0.46	0.28	0.61	0.81
11	Patient anxiety	0.42	0.00	0.53	0.44	0.26	0.59	0.73
12	Family anxiety	0.89	0.00	1.04	0.27	0.01	0.50	1.02
13	Depression	0.84	0.02	1.07	0.52	0.35	0.65	1.04
14	Feeling at peace	0.56	0.00	0.68	0.45	0.25	0.60	0.83
15	Sharing feelings	0.86	0.00	1.14	0.34	0.11	0.50	1.07
16	Information	0.88	0.02	1.25	0.14	0.09	0.35	1.13
17	Practical matters	0.61	0.08	0.92	0.20	0.03	0.42	1.00
	IPOS Total Score	1.86	0.00	2.01	0.64	0.34	0.81	1.42
	IPOS Physical	12.06	38.31	37.53	0.57	0.42	0.69	0.87
	IPOS Emotional	9.99	38.27	35.98	0.45	0.27	0.60	0.86
	IPOS Communication /Practical Issues	16.75	5.46	20.78	0.26	0.05	0.44	0.87

Appendix Table 7: Inter-rater agreement between patient and staff IPOS version at both time points

	Item	No of matched pairs	Patient score (% severe)	Staff score (% severe)	ICC$_a$	Spearman's correlation	Proportion agreement within 1 score
At first assessment (361 matched pairs)							
1	Pain	348	31.9	22.7	**0.59**	0.60	87.1
2	Shortness of breath	345	17.7	17.7	**0.62**	0.63	86.1
3	Weakness or lack of energy	350	50.1	39.3	*0.29*	0.29	82.3
4	Nausea	346	13.9	8.0	**0.46**	0.48	81.2
5	Vomiting	342	7.8	5.0	**0.58**	0.54	88.3
6	Poor appetite	339	25.5	12.7	**0.38**	0.36	74.9
7	Constipation	342	22.7	8.9	**0.47**	0.52	77.5
8	Sore or dry mouth	343	30.5	10.2	*0.25*	0.31	65.1
9	Drowsiness	350	31.9	7.8	*0.11*	0.15	60.6
10	Poor mobility	348	53.7	31.9	**0.42**	0.47	74.4
11	Patient anxiety	347	41.6	37.4	**0.35**	0.36	75.2
12	Family anxiety	283	67.6	41.6	**0.34**	0.37	79.2
13	Depression	348	24.9	18.0	**0.38**	0.39	75.9
14	Feeling at peace	330	25.2	24.1	*0.26*	0.26	72.4
15	Sharing feelings	308	23.5	16.9	*0.13*	0.14	68.8
16	Information	332	15.0	12.7	*0.02*	0.02	70.2
17	Practical matters	317	12.5	7.5	*0.10*	0.10	68.5
At second assessment (260 matched pairs)							
1	Pain	250	21.9	13.5	**0.51**	0.52	86.4
2	Shortness of breath	249	16.2	10.8	**0.52**	0.5	79.1
3	Weakness or lack of energy	255	41.2	30.8	**0.30**	0.31	82.3
4	Nausea	251	9.2	4.2	**0.51**	0.55	86.5
5	Vomiting	250	5.8	2.7	**0.41**	0.43	89.2
6	Poor appetite	245	18.5	10.0	**0.37**	0.41	75.1
7	Constipation	247	17.7	10.8	**0.41**	0.43	74.5
8	Sore or dry mouth	250	26.9	5.0	*0.15*	0.22	60.8
9	Drowsiness	247	27.3	9.6	*0.23*	0.29	66.0
10	Poor mobility	254	47.3	26.5	**0.40**	0.47	70.1
11	Patient anxiety	251	28.5	23.1	**0.36**	0.39	78.5
12	Family anxiety	217	58.1	31.9	*0.27*	0.29	67.7
13	Depression	251	21.2	13.1	**0.40**	0.41	75.7
14	Feeling at peace	243	18.1	22.3	*0.24*	0.25	76.1
15	Sharing feelings	223	20.4	17.3	*0.13*	0.13	67.7
16	Information	246	15.4	8.5	*0.20*	0.21	79.3
17	Practical matters	245	8.8	6.2	*0.10*	0.10	78.4

Correlation was calculated using all answer options.

Appendix: Distribution of items and factor analysis.

A feature of the data reported in the Results affecting results of validity and reliability may be the differing prevalence of symptoms. In the symptom list component of the IPOS, floor effects were present for the symptoms of shortness of breath, nausea, vomiting (73% choosing the 'not at all' response option), constipation, and sore or dry mouth. These low prevalence rates affected the factor solution and made it necessary to parcel the gastrointestinal items (nausea, vomiting, constipation) to achieve adequate fit of the model. On the one hand, these data demonstrate that we succeeded in moving the range and appropriateness of a tool such as the IPOS further upstream in the illness trajectory, sampling from a population with advanced disease but not only at the very end of life. Thus, it may be clinically more representative of the population itself, but the decision to leave these low scoring symptoms in the symptom list is made on clinical and not statistical grounds. Despite some of the symptoms having low loadings, it is important to retain such symptoms as they may indicate severe problems that drive need and require medical care. We therefore do not recommend to further reduce the symptom list on the basis of the factor analytical results, also because of the comparability of models to prior factor analyses run on the core-POS and the APCA African POS [16, 17].

Among the tested solutions, the 3-factor solution performed best. This solution is very similar to the one obtained on the APCA African POS, with the difference of the added symptoms forming one factor and the aspect of sharing feelings with family and friends now clustering with the communication and quality of care items [17]. The item pain, loading on the symptom factor, was the only item with a factor loading below 0.30. This suggests that while most physical symptoms operate together, pain may not readily be collapsed into the construct of 'physical symptoms'. This result warrants further exploration, particularly given the diversity of settings and patients that were included. It may well be explained by underlying heterogeneity in the sample which could be explored by latent mixture modelling [90]. It should also be investigated whether pain forms an overarching factor, affecting and explaining the other factors and subscales in the IPOS.