Chapter 9, assignment 2

8 studies on the reliability of the QBPDS

Table 9.13 Characteristics of included studies

Study	Population	Duration of complaints ± SD			Setting	N	Mean age ± SD (yr)	Male (%)	Country
Schoppink et al. (1)	Chronic LBP	0-6 mo 7-12 mo 13-24 mo >25 mo	14% 6% 8% 72%		General practice	120	39.7 ± 10.4 (R: 21-60)	60	The Netherlands
Rodrigues et al. (2)	LBP	3-6mo 6-12mo 12-18mo 18-24mo >2yr	3.7% 9.2% 5.5% 11.1% 70.3%		Orthopeadics and traumatology clinic	54	44.3 ± 12.1	30	Brazil
Hicks et al. (3)	Current LBP	31.8 ± 23.4 (wk)			4 Continuing care home retirement communities	107	79.6 ± 5.7	28.1	USA
Mousavi et al. (4)	Chronic LBP	6.97 ±8.77 (yr) (R: 0.8-40)			Physical therapy unit of a large hospital	100	40.14 ± 11.59 (R: 17-68)	45	Iran
De Beer et al. (5)	LBP				5 hospitals	100	42 ± 9.13 (R: 23-63)	5	South Africa
Davidson et al. (6)	LBP		UG	IG	Physical therapy outpatient departments of 3 hospitals, 3 community health services and 4 private physical therapy practices	207	UG: 55 ± 17 (R: 19-83) IG: 49 ± 16 (R: 20-80)	UG: 36.2 (n=47) IG: 26.9 (n=52)	Australia
		<1wk 1-6wk 6wk-6mo >6mo	4.2% 21.3% 23.4% 19.2%	17.3% 42.2% 51.1% 17.3%					
Fritz et al. (7)	Acute LBP (<3 wk)	6.2 ± 5.3 (days) (R: 0-19)			Physical therapy	67	39.2 ± 9.7 (R: 21-58)	57	USA
Melikoglu et al. (8)	LBP (>3 wk)	50.89 ± 50.25 (mo)			University medical faculty physical medicine and rehabilitation department	100	45.44 ± 15.05	26	Turkey

(..) = reference number, LBP = low back pain, UG = unchanged group, IG = improved group, FG = first group, SG = second group, N = included population, IQR = interquartile range, R = range, SD = standard deviation, yr = year, wk = week, mo = month, % = percentage

Table 9.14 COSMIN Box Reliability with 4-point rating scale

		excellent	**good**	**fair**	**poor**
Design requirements					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Were at least two measurements available?	Two measurements			Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	Measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated	
7	Were patients stable in the interim period on the construct to be measured?	Patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8	Was the time interval appropriate?	Time interval appropriate (>1 day and <2 weeks)	Time interval reasonable (>2 weeks and < 1 month)	Doubtful whether time interval was appropriate (1 day)	Time interval NOT appropriate (<1day or > 1month)
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there other important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Statistical methods					
11	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated	Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	For ordinal scores: Was a weighted kappa calculated?	Weighted kappa calculated		Unweighted kappa calculated	Only percentage agreement calculated
14	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Table 9.15 Methodological quality of 8 studies on the reliability of the Quebec Pain Disability Scale

		1		**2**		**3**		**4**		**5**		**6**		**7**		**8**	
1	Was the percentage of missing items given?	11%T1 19%T2	e	no	g	no	g	0	e	no	g	no	g	no	g	no	g
2	Was there a description of how missing items were handled?	deleted	e	no	f	deleted	e	NA		no	f	no	f	no	f	no	f
3	Was the sample size included in the analysis adequate?	89	g	54	g	56	g	31	f	31	f	47	f	23	p	100	e
4	Were at least two measurements available?	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e
5	Were the administrations independent?	assumable	g	assumable	g	assumable	g	assumable	g	assumable	g	assumable	g	assumable	g	assumable	g
6	Was the time interval stated? (days)	7	e	3-4	e	11	e	1	e	1	e	42	e	28	e	1	e
7	Were patients stable in the interim period on the construct to be measured?	assumable	g	assumable	g	based on GRS	e	assumable	g	assumable	g	based on GRS	e	based on GRS	e	assumable	g
8	Was the time interval appropriate?	yes	e	yes	e	yes	e	doubtful	f	doubtful	f	no	p	Reasonable	g	doubtful	f
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	2x mail	e	2x by rater	e	2x mail	e	unclear	f	2x clinic	e	2x mail	e	2x clinic	e	2x clinic	e
10	Were there any important flaws in the design or methods of the study?	no	e	no	e	no	e	no	e	no	e	no	e	no	e	no	e
11	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e	yes	e
12	For dichotomous/nominal/ordinal scores: Was kappa calculated?	NA		NA		NA		NA		NA		NA		NA		NA	
13	For ordinal scores: Was a weighted kappa calculated?	NA		NA		NA		NA		NA		NA		NA		NA	
14	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	NA		NA		NA		NA		NA		NA		NA		NA	
	QUALITY SCORE PER BOX	**GOOD**	**g**	**FAIR**	**g**	**GOOD**	**g**	**FAIR**	**f**	**FAIR**	**f**	**POOR**	**f**	**POOR**	**p**	**FAIR**	**f**

NA=not applicable, GRS=Global Rating Scale, e=excellent, g=good, f=fair, p=poor.

Table 9.16 Results of the measurement property reliability (Intraclass Correlation Coefficients (ICC))

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
N	89	54	56	31	31	47	23	100
ICC	0.90	0.93 (0.89-0.96)[1] 0.96 (0.94-0.98)[2]	0.94 (0.90-0.97)	0.86	0.91	0.84 (0.73-0.91)	0.55 (0.20-0.78)	0.92

[1] intra-observer reliability
[2] inter-observer reliability

Table 9.17 Levels of evidence

Level	Rating	Criteria
strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
limited	+ or -	One study of fair methodological quality
conflicting	+/-	Conflicting findings
unknown	?	Only studies of poor methodological quality

+ = positive rating, ? = indeterminate rating, - = negative rating