Book: Measurement in medicine

Authors: HCW de Vet, CB Terwee, LB Mokkink, DL Knol

www.cambridge.org/9780521133852

Answers Chapter 5

1. Calculation and interpretation of ICC (see details assignment 1)

1.A

	Mary (M) Mean (SD)	Peter (P) Mean (SD)	Mean M-P (SD_{M-P})	$ICC_{agreement}$ 95% CI	$ICC_{consistency}$ 95% CI
Affected side	69.5 (17.6)	68.7 (16.2)	0.8 (10.0)	0.83 (0.77 – 0.87)	0.83 (0.77 – 0.87)
Non-affected side	79.8 (7.6)	78.9 (8.4)	0.9 (9.6)	0.28 (0.13 – 0.42)	0.28 (0.13 – 0.42)

1.B Although the values of $ICC_{consistency}$ and $ICC_{agreement}$ are the same, the agreement parameter is preferred, because it takes the systematic error into account. In this example there is hardly any systematic error, but if you had not calculated the $ICC_{agreement}$, you would not have known. The VARCOMP analysis is preferred to obtain the values for each variance component. The Reliability analysis is necessary to obtain 95% confidence intervals.

1.C The ICC for the affected side is much higher than the ICC for the non-affected side. This might be due to greater measurement error for the non-affected side. However, there is no reason why this should be the case. If the measurement errors were different, the affected side is expected to give the greatest measurement error, because depending on the pain, the patient might stop the movement earlier or later. Another explanation is that the variation in the patients differs between the affected side and the non-affected side. This is a plausible explanation. We can see in the above table that the SD of the measurements of the affected side (17.6 for Mary and 16.2 for Peter) is much greater than the SD for

the non-affected side (7.6 and 8.4, respectively). If the scores are homogeneous, it is more difficult to distinguish between the patients.

2. Calculation of measurement error (use the output obtained by assignment 1)

2.A $SEM_{agreement}$ for the example concerning shoulder range of movement (ROM) can be calculated by taking the square root of the error variance of the ICC formula, using the variance components from the VARCOMP analysis:

Affected side: $SEM_{agreement} = \sqrt{(\sigma_o^2 + \sigma_{residual}^2)} = \sqrt{(0.0 + 49.981)} = 7.0697$

Non-affected side: $SEM_{agreement} = \sqrt{(\sigma_o^2 + \sigma_{residual}^2)} = \sqrt{(0.106 + 45.910)} = \sqrt{46.016} = 6.7835$

2.A $SEM_{consistency}$ can be calculated in the three ways that are presented in section 5.4.2.1.

1. by taking the square root of the error variance of the ICC formula (using the variance components from the VARCOMP analysis:

Affected side: $SEM_{consistency} = \sqrt{\sigma_{residual}^2} = \sqrt{49.981} = 7.0697$

Non-affected side: $SEM_{consistency} = \sqrt{\sigma_{residual}^2} = \sqrt{45.910} = 6.7757$

2. by using the formula: $SD_{difference} / \sqrt{2}$:

Affected side: $SEM_{consistency} = SD_{difference} / \sqrt{2} = 9.9983 / 1.41421 = 7.0699$

Non-affected side: $SEM_{consistency} = SD_{difference} / \sqrt{2} = 9.5823 / 1.41421 = 6.7757$

3. by using the formula: $SEM_{consistency} = \sigma_y \sqrt{(1 - ICC_{consistency})}$:

For the affected side, we take the SD_{pooled} as σ_y:

$$\sqrt{\frac{SD_1^2 + SD_2^2}{2}} = \sqrt{\frac{17.604^2 + 16.246^2}{2}} = 16.9386$$

and the $ICC_{consistency} = 0.8258$

$SEM = SD_{pooled}\sqrt{(1 - ICC_{consistency})} = 16.9386 \times \sqrt{(1 - 0.8258)} = 7.0697$

For the non-affected side, we take the SD_{pooled} as σ_y :

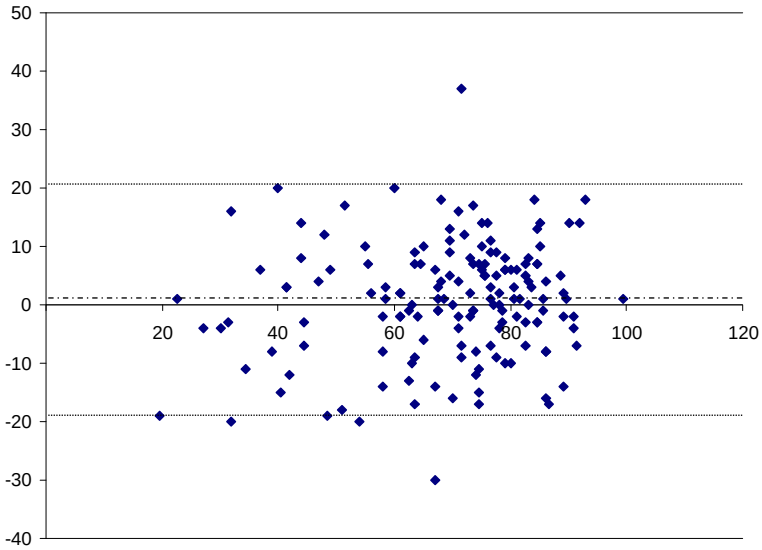
$$\sqrt{\frac{SD_1^2 + SD_2^2}{2}} = \sqrt{\frac{7.600^2 + 8.380^2}{2}} = 7.9995$$

and the $ICC_{consistency} = 0.2825$

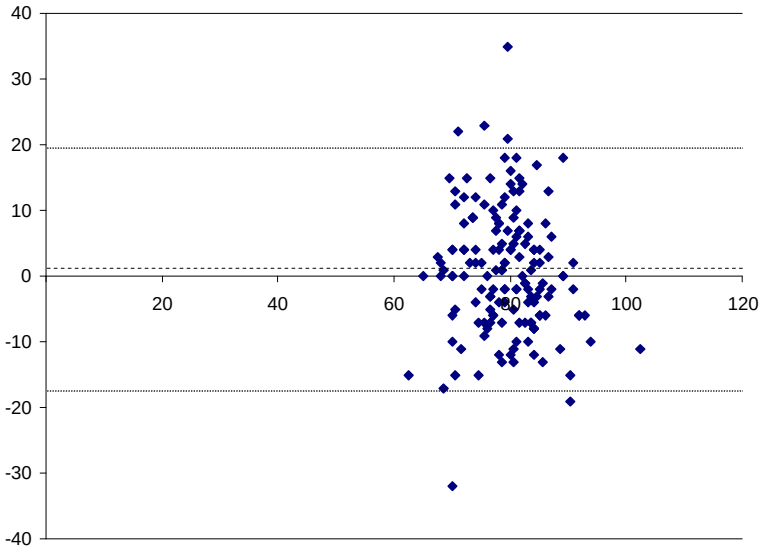
$SEM_{consistency} = SD_{pooled}\sqrt{(1 - ICC_{consistency})} = 7.9995 \times \sqrt{(1 - 0.2825)} = 6.7760$

2.B The affected side and non-affected side have exactly the same SEM values (both $SEM_{agreement}$ and $SEM_{consistency}$). This implies that the difference in ICC values is totally due to variation between patients in the ROM of the affected shoulder and the non-affected shoulder.

2.C Bland and Altman plot



Affected shoulder:
$\bar{d}$ = 0.80
Limits of agreement:
Lower limit: -18.80
Higher limit: +20.40



Non affected shoulder
$\bar{d}$ = 0.90
Limits of agreement:
Lower limit: -17.91
Higher limit: +19.70

2.D Bland and Altman plot with 95% CI of limits of agreement

The limits of agreement are defined as the mean difference between the physiotherapists ± 1.96 × $SD_{difference}$

Affected side:

Lower limit: 0.8000 – 1.96 × 9.9983 = -18.7967; upper limit: 0.8000 + 1.96 × 9.9983 = 20.3967

Non-affected side:

Lower limit: 0.8968 – 1.96 × 9.5823 = -17.8845; upper limit: 0.8968 + 1.96 × 9.5823 = 19.6781

The 95% of the Limits of agreement are: Limits of agreement ± 1.96 × √3 × $SD_{difference}$ / √*n*

* Affected shoulder:

Lower limit: -18.7962 – 1.96 × √3 × 9.998 / √155 = -18.796 – 1.96 × 0.139 = -19.069

Lower limit: -18.7962 + 1.96 × √3 × 9.998 / √155 = -18.796 + 1.96 × 0.139 = -18.524

Upper limit: 20.3962 – 1.96 × √3 × 9.998 / √155 = 20.396 – 1.96 × 0.139 = 20.124

Upper limit: 20.3962 + 1.96 × √3 × 9.998 / √155 = 20.396 + 1.96 × 0.139 = 20.668

* Non-affected shoulder:

Lower limit: -17.9061 – 1.96 × √3 × 9.582 / √155 = -17.9061 – 1.96 × 1.333 = -20.519

Lower limit: -17.9061 + 1.96 × √3 × 9.582 / √155 = -17.9061 + 1.96 × 1.333 = -15.293

Upper limit: 19.6997 – 1.96 × √3 × 9.582 / √155 = 19.6997 – 1.96 × 1.333 = 17.087

Upper limit: 19.6997 + 1.96 × √3 × 9.582 / √155 = 19.6997 + 1.96 × 1.333 = 22.313

3. Calculation of SEM by rewriting the ICC formula

Note that this SEM value of 3.30 is much smaller than the SEM value found for the non-affected side, i.e. 6.78. The SEM value is under-estimated, because the ICC value that we used in our formula only holds for very heterogeneous populations. For the homogeneous ROM values for the non-affected shoulder (with a pooled SD of 8.00) the ICC value of 0.28 holds.

4. Calculation of kappa

4.A

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

$$P_o = \frac{17+8}{30} = 0.833$$

$$P_o = \frac{17}{30} \times \frac{22}{30} + \frac{13}{30} \times \frac{8}{30} = 0.531$$

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.833 - 0.531}{1 - 0.531} = 0.645$$

4.B

We used the program available at: http://faculty.vassar.edu/lowry/kappa.html

The computer programme produces a kappa value of 0.645 (95% CI: 0.360 – 0.929)

Unweighted Kappa			
		.95 Confidence Interval	
Observed Kappa	Standard Error	Lower Limit	Upper Limit
0.6445	0.1451	0.3601	0.9289

0.6445	maximum possible unweighted kappa, given the observed marginal frequencies
1	observed as proportion of maximum possible

4.C When we consider the schemes developed by Landis and Koch (1977) and by Fleiss (1981), the value falls within the category ‘substantial’ and ‘fair to good’, respectively. However, there were only 30 EEGs in this study, and from the 95% CI we can conclude that the kappa may also take values which correspond to excellent (kappa > 0.80) and poor (kappa < 0.40) agreement. The numbers are therefore too small to draw a firm conclusion.

		C.I.S.	CIN 3	CIN 2	CIN 1	ND	Total
C.I.S.	O E W	1 0.022 0	0 0.226 1	0 0.409 4	0 0.301 9	0 0.043 16	1
CIN 3	O E W	1 0.516 1	13 5.419 0	9 9.806 1	1 7.226 4	0 1.032 9	24
CIN 2	O E W	0 0.731 4	7 7.677 1	18 13.892 0	9 10.237 1	0 1.462 4	34
CIN 1	O E W	0 0.624 9	1 6.548 4	11 11.849 1	15 8.731 0	2 1.247 1	29
ND	O E W	0 0.108 16	0 1.129 9	0 2.043 4	3 1.505 1	2 0.215 0	5
Total		2	21	38	28	4	93

5. Calculation of weighted kappa value

In the table the top row in each cell represents the observed agreement (O), followed by the expected agreement (E) on the second row, and the quadratic weight (W) on the third row.

$$\kappa = 1 - \frac{\sum w_{ij} \times P_{o_{ij}}}{\sum w_{ij} \times P_{e_{ij}}} =$$

$$1 - \frac{0\times1 + 1\times0 + 4\times0 + \ldots + 4\times0 + 1\times3 + 0\times2}{0\times0.022 + 1\times0.226 + 4\times0.409 + \ldots + 4\times2.043 + 1\times1.505 + 0\times0.215} = 1 - \frac{50}{146.929} = 0.660$$

$\sum w_{ij} \times P_{o_{ij}}$ = (from first row left to last row right: weight × observed agreement) = 50

$\sum w_{ij} \times P_{e_{ij}}$ = (from first row left to last row right: weight × expected agreement) = 146.929

6. Design of a Generalisability study

Researchers developed the Pain Assessment Checklist for Seniors with Limited Ability to Communicate, an observation scale for the assessment of pain in elderly people with dementia. Nursing home doctors decide to introduce this scale in their nursing homes.

6.A Potential sources of variation and related "Decision" question:

1. Minute to minute variation: how many minutes should the observation period last?

2. Hour to hour variation: how many times a day should they measure?

3. Day to day variation: should they measure on different days, and on how many days?

4. Observation before or after a meal: does timing over the day depend on meals?

5. Observation by nursing home doctor, by nurse or by family member: does it make a difference who observes?

6.B Measurement scheme of 50 patients with dementia, measurements made by 2 nursing home doctors, during three different periods of a day (T1, T2, T3), and during each period observing three times for 2 minutes (1, 2, 3). This Table concerns Day 1.

Day 1	Nursing home doctor 1									Nursing home doctor 2								
	T1			T2			T3			T1			T2			T3		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1																		
2																		
•																		
•																		
49																		
50																		

7. Generalisabilty and Decision studies

7.A $\sigma^2_p = 70$; $\sigma^2_{po} = 30$; $\sigma^2_{pm} = 12$; and $\sigma^2_{residual} = 15$

The basic formula is G consistency is

$$G_{consistency} = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{po} + \sigma^2_{pm} + \sigma^2_{residual}} = \frac{70}{70 + 30 + 12 + 15} = 0.551$$

1 measurement by 2 different clinicians (dividing variance components containing '*o*' by 2) leads to:

$$G_{consistency} = \frac{\sigma^2_p}{\sigma^2_p + \frac{\sigma^2_{po}}{2} + \sigma^2_{pm} + \frac{\sigma^2_{residual}}{2}} = \frac{70}{70 + \frac{30}{2} + 12 + \frac{15}{2}} = 0.670$$

And 5 measurements by 1 clinician (dividing variance components containing '*m*' by 5) leads to:

$$G_{consistency} = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{po} + \frac{\sigma^2_{pm}}{5} + \frac{\sigma^2_{residual}}{5}} = \frac{70}{70 + 30 + \frac{12}{5} + \frac{15}{5}} = 0.664$$

Decision: Comparing 0.670 to 0.664, measurements made by two different clinicians is slightly more reliable.

7.B We had ignored systematic errors, but if there are systematic errors, we expect larger errors between clinicians, because they may use different techniques or different thresholds to decide when they hear a blood flow, while one clinician is often more consistent in his or her technique and threshold. By incorporating the systematic errors, 5 measurements made by one clinician is probably the (slightly) preferred strategy.