Book: Measurement in medicine

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

www.cambridge.org/9780521133852

Answers Chapter 8

1. Distributions

A. The Table shows the mean scores and standard deviations of three relevant patient populations. CFS patients have average scores between 11 and 16 on the various domains, chronically unwell patients have average scores between 9 and 13, and well patients score on average between 6.5 and 8.5. For all three groups the domains 'reduced activity' and 'reduced motivation' have the lowest mean scores, and the domain 'general fatigue' has the highest mean score.

Further, it can be seen that the whole range of the scale is used by these three groups. This can be seen by looking at the range of scores for each domain, but also by looking at the borders of the upper and lower quartile.

B. The presentation of medians and quartiles is quite informative. The 'distance' between the quartiles and the medians give an indication of the skewness of the data. The same information can be deduced from comparison of mean and median values. A histogram for each domain in each group would have been even more informative, but journals seldom allow this, and it adds little to the current information.

C. Ceiling effects do not occur. In the CFS group 13% of the population has the maximum score of 20. In the study sample consisting of 'well patients' there were three domains with about 25% of the patients having the lowest score. As we explained in Section 8.3.4 if these patients do not suffer from fatigue in any way, we do not label this as floor effects. Whether there are floor effects depends on the aim of the study. When the MFI-20 is used to evaluate an intervention to reduce fatigue in the general population, for example by increasing the hours of bed rest, there might be a floor effect, because this population is already scoring low on various domains of the MFI. In that case an instrument which especially focuses on mild fatigue might be more appropriate.

2. Anchor-based MIC distribution

We follow the steps as described in Section 8.5.4.3.

A. (Step 1) We first make a Table to present the mean change values (T0 – T2) for the various categories on the anchor. This is the same table as Table 8.3 in Section 8.5.4.3, except for the last lines.

Table 8.7 The mean change (T0 – T2) of the PRAFAB questionnaire scores by global rating scale (GRS9) at 12 weeks follow-up for the total group of patients with stress urinary incontinence

	Number of patients	Change in PRAFAB score
Global rating scale	N = 534	$Mean_{change}$ (SD_{change})
1 Completely recovered	124	6.51 (1.84)
2 Much improved	86	4.52 (1.71)
3 Moderately improved	86	3.57 (1.33)
4 Slightly improved	49	2.55 (0.79)
5 Not changed	139	0.82 (0.98)
6 Slightly deteriorated	39	-0.36 (1.06)
7 Moderately deteriorated	7	-2.29 (0.76)
8 Much deteriorated	3	-4.00 (1.73)
9 Worse than ever	1	-6.00 (---)
Importantly improved (1,2,3,4)	345	4.72 (2.15)
Not importantly changed (5)	139	0.82 (0.98)
Importantly deteriorated (6,7,8,9)	50	- 0.96 (1.64)

B. (Step 2) To obtain the MIC value when the cut-off point on the anchor is laid between the categories ‘no change’ and ‘slightly improved’, we divide the population in two parts: there are 345 patients who importantly improved and 139 patients who did not change. The distribution of the changes scores on the PRAFAB for both samples are calculated in SPSS (see www.clinimetrics.nl) and presented in Table 8.8.

Table 8.8 Change scores on PRAFAB of "importantly improved" and "not importantly improved" groups (T0 – T2), and corresponding values for sensitivity and specificity. Cut-off point on anchor between slightly improved and not changed.

Change score T0-T2	**"Importantly improved" group N**	**proportion**	**"Not importantly changed" group N**	**proportion**	**ROC cut-off point on PRAFAB**	**Sens**	**Spec**	**1-sens**	**1-spec**	**[1-sens] + [1-spec]**
11	0	0	0	0	11.00	0.000	1.000	1.000	0.000	1.000
10	4	0.012	0	0	9.50	0.012	1.000	0.988	0.000	0.988
9	16	0.046	0	0	8.50	0.058	1.000	0.942	0.000	0.942
8	25	0.072	0	0	7.50	0.130	1.000	0.870	0.000	0.870
7	31	0.090	0	0	6.50	0.220	1.000	0.780	0.000	0.780
6	43	0.125	0	0	5.50	0.345	1.000	0.655	0.000	0.665
5	52	0.151	0	0	4.50	0.496	1.000	0.504	0.000	0.504
4	52	0.151	1	0.007	3.50	0.646	0.993	0.354	0.007	0.361
3	59	0.171	4	0.029	**2.50**	**0.817**	**0.964**	**0.183**	**0.036**	**0.219**
2	61	0.177	25	0.180	1.50	0.994	0.784	0.006	0.216	0.222
1	2	0.006	58	0.417	0.50	1.000	0.367	0.000	0.633	0.633
0	0	0	44	0.317	-0.50	1.000	0.050	0.000	0.950	0.950
-1	0	0	4	0.029	-1.50	1.000	0.021	0.000	0.979	0.979
-2	0	0	3	0.021	-2.50	1.000	0.000	0.000	1.000	1.000
-3	0	0	0	0	-3.50	1.000	0.000	0.000	1.000	1.000
-4	0	0	0	0	-4.50	1.000	0.000	0.000	1.000	1.000
total	345	1	139	1						

The optimal ROC cut-off point = 2.5

Column 1 represents the change scores (variable changeT0_T2): totalPRAFABT0 - totalPRAFABT0.

Columns 2 and 4 represent the absolute distribution of the 'importantly improved' and 'non-changed' groups, respectively.

Columns 3 and 5 represent the relative distribution of the 'importantly improved' and 'non-changed' groups, respectively.

Column 6 represents the ROC cut-off points. The SPSS program uses 0.5 scores as the cut-off points. Although only integers occur as change scores, these 0.5 scores are very clear as cut- off points. If the optimal score is 2.5 and we take that as MIC value, then it is clear that a change score of 2 is below the MIC and 3 is above the MIC.

Column 7 represents the sensitivity, which is calculated as the fraction of the persons, importantly improved according to the anchor, who have change scores as large as the cut-off point. For example at the cut-off point of 2.5 the sensitivity is 0.817. Note that this column is the cumulative summation of column 3, starting from the top.

Column 8 represents the specificity, which is calculated as the fraction of the persons, not changed according to the anchor, who have change scores smaller than the cut-off point. For example at the cut-off point of 2.5 the specificity is 0.964. Note that this column is the cumulative summation of column 5, starting from the bottom.

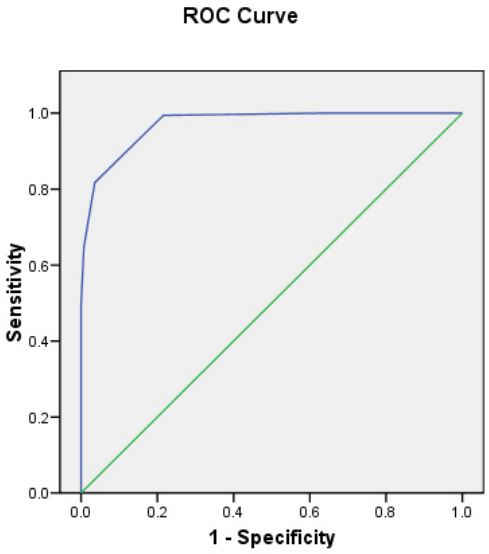
Columns 9 and 10 are the complements of Columns 7 and 8, and represent the percentages of false negative and false positive misclassifications, respectively.

Column 11 represents the sum of columns 9 and 10, i.e. the sum of the percentages of misclassifications.

C. (Step 3) The findings in columns 7 and 10 are obtained by ROC analysis which presents the sensitivity and (1 – specificity), when choosing the option /PRINT = SE COORDINATES.

In Table 8.6 we see that the optimal cut-off point is 2.5, but the sum of percentages misclassification (0.219) is very close to the sum of percentages misclassification for the cut-off point 1.5 (0.222). This can also be seen in the ROC plot by the two angles in the ROC curve. From the data in Table 8.6, and from the ROC output (Figure and Table), it can be seen that cut-off 2.5 is the one with sensitivity about 0.80 and specificity of about 0.95. The cut-off point 1.5 is the one with higher sensitivity (0.99) and some what lower specificity (0.78).

Positive if Greater Than or Equal To(a)	Sensitivity	1 - Specificity
-3.0000	1.000	1.000
-1.5000	1.000	.978
-.5000	1.000	.950
.5000	1.000	.633
1.5000	.994	.216
2.5000	.817	.036
3.5000	.646	.007
4.5000	.496	.000
5.5000	.345	.000
6.5000	.220	.000
7.5000	.130	.000
8.5000	.058	.000
9.5000	.012	.000
11.0000	.000	.000



Coordinates of the Curve

Figure 8.13 ROC curve representing the sensitivity and 1- specificity at various change scores on the PRAFAB questionnaire. Test Result Variable(s): changeT0_T2 . a The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values. Based on Hendriks et al. (2008), with permission.

D. These data can be imported in an excel file to obtain a Figure showing the relative frequency distributions of the ‘improved’ and ‘non-changed’ samples. Giving the relative frequency distributions of the improved group a ‘minus’ sign results in a curve on the left side of the X-axis for the improved group.

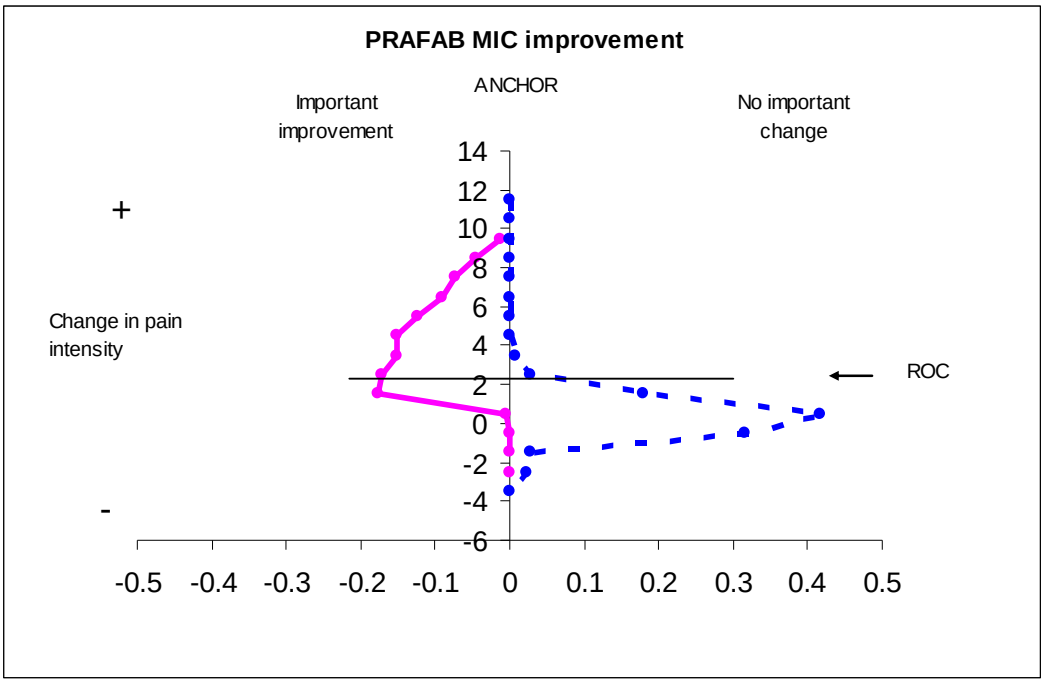


Figure 8.14 Distribution of change scores on the PRAFAB questionnaire of patients who reported an important improvement (n=345) compared to those with no important improvement (n=139) on the anchor (GRS). Based on Hendriks et al. (2008), with permission.

3. Response shift

A. This Table provides information about reconceptualisation. It shows that the frequencies of the nominated areas have changed, implying that at least part of the persons has changed some of their five life areas.

B. From the way the data are presented, it is difficult to see how many persons changed their five important areas of life. In addition, individual changes may get unobserved, e.g. if one person replaces work by social life and another person changes the other way around, this will not appear in this table. A better insight in how often the life areas were changed could have been provided by presenting the percentage of people who did not change, who changed one area, two areas and so on.

C. This statement concerns the relative weights assigned to each of the 5 areas of life, adding up to 100% and represents reprioritization of what is most important in life.

D. Although the ICC can be used for this purpose, it does not provide a good overview of the magnitude of changes that occurred (in the same way that the ICC for reliability does not give much information about the magnitude of the measurement error).
It is difficult to decide how to handle these weights if patients made a change in their 5 important areas for the post test, i.e. reconceptualisation occurred. The authors could have restricted this analysis to the persons who did not change their five important areas of life. Then it would have been possible to present the mean value and standard deviations or range of absolute changes on the areas.

E. The patients are instructed to consider the same areas of importance and then rate their pre-test situation using the post-test perceptions: "show me how you now think you were doing". The difference between results of the pre-test and then-test points to recalibration.

F. Without taking response shift into account the effect of the rehabilitation program is 5.09. With taking response shift into account, i.e. adding the difference of the pre-test and then test score (the response shift effect) to the difference between post-test and pre-test means that the total effect of the rehabilitation program would amount to + 5.09 + 9.56 = 14.65. (see also Figure 8.15)

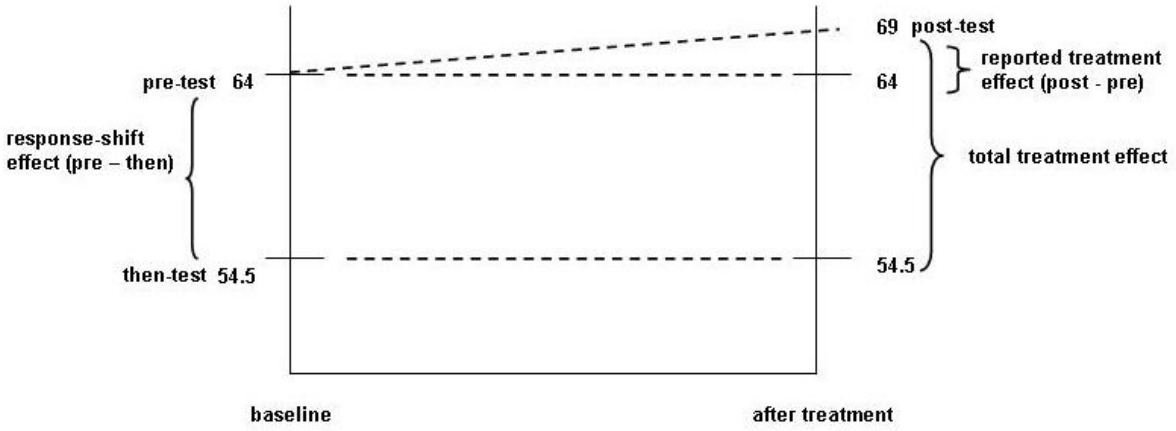


Figure 8.15 Pre-test, post-test and then test scores on the SEIQOL-DW. Based on Dempster et al. (2010), with permission.

G. The response shift effect might be caused by recalibration, reconceptualisation and reprioritization. Note that in question E the instruction was formulated in such a way that only recalibration was measured. The difference between pre-test and post-test scores calculated as ∑relative weight x VAS score, also includes the effects of reconceptualisation and reprioritization, and the contribution of the three mechanisms cannot be disentangled.

H. Whether you label the response shift effect as bias depends on which effects of the rehabilitation program you are interested in. There is no bias if you are really interested in the HRQL as perceived by the patients at the moments of measurements. In that case you allow the definition of HRQL to be changed over time. There is bias, if you are interested in whether the HRQL as perceived during the pre-test assessment would improve. In that case you are interested in the total treatment effect. The observed difference between pre-test and post-test would give a biased estimate of this effect.

However, the authors did not describe what the aim of the rehabilitation programme was and how they expected the programme to change HRQL. Often rehabilitation programmes are not only aimed at improvement of the physical fitness of patients, but also aim to increase reassurance and self confidence, and reduce anxiety and depression. Moreover, during rehabilitation programmes patients often learn to set new goals in life and reconsider their possibilities. In that case the aim of the rehabilitation program is to induce response shift. Considering the aim of most rehabilitation programs it is tempting to say that HRQL has improved by 14.65 points instead of 5.09 points.

A few additional remarks:

- A more qualitative description of what happened would have been worthwhile.
- This is a before-after assessment. Without a control group it is unclear how much of the changes can be attributed to the rehabilitation programme above the natural course.

Note: This assignment is based on the model and mechanisms proposed by Schwarz and Sprangers (1999). The ideas about response shift and its mechanisms are still in development and new publications present adaptations and refinements of the model.