REPRODUCIBILITY OF RANGE OF MOTION MEASUREMENTS OF THE SPINE WITH THE CYBEX EDI 320

Bart Koes, Henk v Mameren, Lex Bouter, Alex Essers, Willem Elzinga, Gard Verstegen and Fons Kessels.

University of Limburg, Department of Epidemiology and Health Care Research, P.O. Box 616, 6200 MD Maastricht, the Netherlands

Range of motion (ROM) measurements of joints are often used in physiotherapy for diagnosis and for the selection of an appropriate intervention. In addition, it makes it feasible to quantify the effect of the intervention on the ROM. In this study we investigate the reproducibility (precision or reliability) of ROM measurements of the spine with the CYBEX Electronic Digital Inclinometer (EDI) 320. A further aim of the study is to estimate the minimum effect that can be detected in a clinical trial as a function of group size. In other words, this article deals with random variability in ROM measurements of the spine with the EDI 320. As is explained elsewhere in this volume, lack of random error is one of the main desiderata for an effect parameter (Bouter et al., 1990). We considered reproducibility as the degree in which repeated measurements on the same subjects provide the same results. This can be disturbed by intra-individual variability and measurement error.

METHODS

Reproducibility is determined by comparing repeated measurements on the same subjects (test-retest method). Sixteen healthy subjects (physiotherapy students) participated in the study. ROM measurements were carried out for several spinal movements (table 1).

Table 1: Movements between extremes of the path of motion* and position of the inclinometer

Movement	position inclinometer	position subject
1 cervical forward flexion	os nasale	sitting
2 cervical lateral flexion to the left	os temporale proc. zygomaticus	sitting
3 cervical lateral flexion to the right	os temporale proc. zygomaticus	sitting
4 thoracic forward flexion	T1	standing
5 lumbar forward flexion	T12	standing

*movements 4 and 5:from anatomical to extreme position

These movements were chosen because they are part of the research protocol of a clinical trial which has been conducted recently (Koes et al, 1990). All measurements were carried out by one physiotherapist who was experienced in the use of the inclinometer. Special attention was paid to standardisation of the measurement protocol. This meant that all subjects received the same information concerning the movements to be executed.

Table 2 presents the eight series of measurements which were taken on each subject. Reproducibility of the measurements was determined for the following time intervals:
- consecutive ratings;
- on the same day with a time interval of at least one hour;
- after two days.

Intervals were chosen because of the considerable degree of intra-individual variability known in the cervical spine (Mameren, 1988). Furthermore it presents how the measurements for each time interval were compared. When comparing measurements with one hour and two days time difference the mean of the measurements which were taken consecutively is used.

Table 2. Comparison of the measurements for each time interval

time interval	moment* day 1	day 2	comparison
consecutive ratings	A B X X X X	C D X X	for each moment the difference between two easurements (X,X) is calculated
1 hour	A <-----> B C <-----> D		the difference is calculated between the moments (A and B), and (C and D)
2 days	A <----> C B <----> D		The difference is calculated between (A and C) (A and D), (B and C) and (B and D)

* At the moments A, B, C and D the two measurements (X,X) were carried out consecutively. Between the moments on one day (A and B) and (C and D) there was a time interval of at least one hour.

** Between day 1 and day 2 there was a time interval of two days.

RESULTS

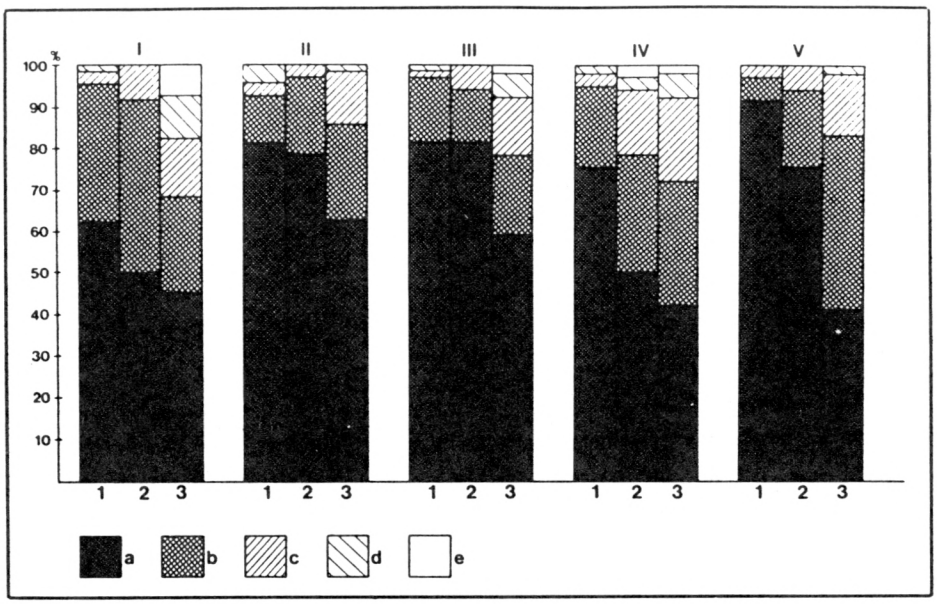
Figure 1 presents the frequency (%) of differences in degrees between the first and second measurement for each of the three time intervals.

For measurements carried out consecutively it appears that in about 90% of the measurements the differences are less than or equal to 10°. What is notable are the differences in results for the category 0° - 5° of lumbar forward flexion (91%) and cervical forward flexion (62%). The latter shows a smaller reproducibility.

For measurements carried out on the same day with a time interval of one hour, it appears that, except for thoracic forward flexion (80%), the differences are less than or equal to 10°, in about 90% of the measurements. For cervical forward flexion and thoracic forward flexion 50% of the measurements carried out with one hour difference fall in the category 0° - 5°. For thoracic forward flexion some differences (4%) were more than 20°.

For measurements carried out with a time interval of two days, it appears that for cervical forward flexion and thoracic forward flexion in about 70% of the measurements the differences are less than or equal to 10°. Differences of more than 20° or more are found for cervical forward flexion (8%), cervical lateral flexion to the right (3%) and thoracic forward flexion (3%). For all ROM it appears that, with an interval of two days, the percentage of measurements with differences less than or equal to 5° is lower than for smaller time intervals between measurements.

Figure 1. Results of the measurements *



* movements:
I : cervical forward flexion
II : cervical lateral flexion to the left
III : cervical lateral flexion to the right
IV : thoracic forward flexion
V : lumbar forward flexion

time intervals:
1 : consecutive ratings
2 : minimum of one hour
3 : two days

frequency (%) of differences between first and second measurement in categories of 5°:
a : 0° - 5°
b : 6° - 10°
c : 11° - 15°
d : 16° - 20°
e : more than 20°

Table 3 presents a calculation to determine which difference in ROM is detectable when comparing groups in a randomized clinical trial (RCT) (Bouter et al, 1990). A 'sample size formula' is used to calculate the minimum number of subjects which are needed in an RCT to detect a difference between the study groups of a certain magnitude (Meinert 1986). However, when the size of the study groups is decided in advance (in the example the experimental group (nt) and the control group (nc) comprise 60 subjects), it is possible to calculate the magnitude of the treatment minimal difference between the groups which can be detected in the study. Components necessary for the calculation of the sample size are:

1. magnitude of the effect (δ A) in which one is interested.
2. acceptable level of type I error (α): probability of finding a significant difference between groups when in reality there is no difference.
3. acceptable level of type II error (β): probability of not finding a significant difference when in reality there is a difference.
4. the variance (σ^2) of the outcome measure. An approximation for this is the mean variance calculated over all moments (A, B, C and D). We assume that the variance (due to random measurement error and intra-individual variability) which occurs in the hypothetical group of 60 subjects would be the same as among the 16 subjects in

our study. Additional variance which could occur because patients may have a different reaction to therapy or show differences in the natural course of their complaints, is assumed not to be present. This may lead to an underestimation of the true variance in an RCT. Results of the calculations must therefore be interpreted with caution.

Table 3. calculation* of the minimal difference in cervical forward flexion which is detectable with study groups of 60 subjects.

subject	measurement A	B	C	D	mean	variance (σ^2)
1	139	127	123	122	127.8	60.9
2	135	140	140	139	138.5	5.7
3	149	141	142	140	143.0	16.7
4	162	153	162	158	158.8	18.3
5	152	148	141	141	145.5	29.7
6	144	151	145	145	146.3	10.3
7	158	157	154	145	153.5	35.0
8	150	150	142	152	148.5	19.7
9	149	146	145	138	144.5	21.7
10	148	163	165	162	159.5	60.3
11	149	141	140	146	144.0	18.0
12	147	145	129	120	135.3	168.3
13	138	128	165	158	147.3	295.6
14	157	158	153	149	154.3	16.9
15	152	155	140	149	149.0	42.0
16	142	132	129	140	135.8	38.9

* sample-size formula

$$\delta A = \sqrt{\frac{2(Z\alpha + Z\beta)^2 \times 2\sigma^2}{n_t}}$$

σ^2 = 858/16 = 53.6
α = 0.05 ($Z\alpha$ = 1.96)
β = 0.1 ($Z\beta$ = 1.282)
n_t = 60

thus: $\delta A = \sqrt{\frac{2(1.96 + 1.282)^2 \; 2 \times 53.6}{60}} = 6.1°$

Table 4 presents the minimum difference (in degrees) for each movement which can be detected in an RCT with two groups of 60 subjects. It appears that differences between the experimental and the control group of 5° (or more) will be detectable; that is, (transversal) differences between two study groups (experimental group and control group).

Table 4. Magnitude of the minimal difference in ROM detectable with study groups of 60 subjects:

cervical forward flexion	6°
cervical lateral flexion to the left	4°
cervical lateral flexion to the right	4°
thoracic forward flexion	5°
lumbar forward flexion	4°

DISCUSSION

The most important results of our study can be summarised as follows:

1. In about 70% of the measurements (for the three different time intervals studied) the first and second measurements differ by less than 10°. For nearly all measurements these differences vary between the 0° and 20°.

2. For all ROM it shows that the longer the time interval the more frequently the differences between the first and second measurements are larger than 5°.
3. Cervical forward flexion provides the least reproducible results, while the lumbar forward flexion measurements are the most reproducible.
4. With sample-size formulas it can be calculated that in an RCT with study groups of 60 subjects differences between the experimental and the control groups of about 5° will be detectable.

All measurements were carried out on subjects without complaints. Therefore conclusions about reproducibility of ROM for patients can not be drawn with certainty. In addition to this, all measurements were carried out by the same physiotherapist who had practised extensively in advance with the inclinometer. It is not known whether another physiotherapist would have performed with the same degree of reproducibility. The longest time interval between the measurements was two days. Strictly speaking it is not possible to extrapolate our results to longer time intervals (e.g. weeks or months).

The results of this study show that ROM cannot be measured with the same reproducibility for different movements. The data for cervical forward flexion were less reproducible than the measurements of the lumbar forward flexion. When measurements were carried out consecutively, it turned out that differences of more than 5° were found for cervical forward flexion in 38% of the measurements and in only 9 % of the measurements of the lumbar forward flexion. The reason for these differences in reproducibility is that for some movements there is greater opportunity for intra-individual variability and/or random measurement error. The following determinants of reproducibility are important.
First, standardization of the measurement protocol of the specific movement is often difficult to realise, for example, due to problems with the positioning of the inclinometer on the specific body segment.
Second, variability of some movements might be larger than for others. For cervical forward flexion this might be the case because motion is possible in more planes than for movements such as lumbar forward flexion.
Third, when the absolute difference between the first and second measurement was determined for each movement, we did not take into account the magnitude of the total ROM of that movement. It is possible that for movements with a larger (total) ROM (for example 150°), the variance of the measurements will be greater than for movements with a smaller (total) ROM of 60°.

What conclusions can be drawn from these data for clinical use?
When measurements are taken on individual patients, one needs to consider that the value which is measured can deviate from the true value because of random measurement error. When after some time measurements are taken again on the same patient, a difference with the first rating can be explained by measurement error, intra-individual variability, the natural course of the disease and a therapeutic effect. In this study with healthy subjects we found differences which vary in most cases between 0° and 20°. Essentially these differences can be attributed to 'physiological' intra-individual variability and random measurement error. So, if we want to draw conclusions for an individual patient about the effect (increase of ROM) of a therapeutic intervention, the difference between the two measurements (before and after the intervention) must be at least 20° to be attributable to the intervention. The assumption is that patients will have at least the same degree of 'physiological' variability as healthy subjects. However, in our study in 70% of the measurements the differences were less than or equal to 10°. Therefore, with some caution differences between 10 and 20° can probably be interpreted as an indication of a therapeutic effect.

When measurements are taken on groups of patients, random measurement error and intra-individual variability play a role in the interpretation of the results as well. However, if we want to study the effect of an intervention (increase in ROM) by comparing an experimental to a control group in an RCT, we can detect differences of less than 20°. The reason for this is that it is not very likely that all patients in one group will show differences of 20° solely because of measurement error and/or intra-individual variability. Furthermore, overestimations and underestimations on the individual level will (partly) cancel each other out and thus the true mean ROM will be approximated better by increasing the size of the study groups. In this study we calculated on the basis of the results from 16 healthy subjects that with study groups

of 60 subjects true differences of about 5° would seem to be detectable.
Another issue is the clinical relevance of an effect of such a magnitude. Does a difference of 5°, 10° or 20° have to be considered relevant? Furthermore, measurement of maximum ROM does not give any information about the quality or the velocity of the movement. But these questions go beyond the reproducibility of the ROM as measured by the EDI 320 and thus fall outside the scope of this article.

References

Bouter LM, Linden van der S and Koes B (1990), How to assess the effects of physiotherapy, (in this volume)

Koes BW, Bouter LM, Knipschild PG, Mameren van H, Essers A, Houben JP, Verstegen GMJR and Hofhuizen DM (1990), The effectiveness of manual therapy, physiotherapy and treatment by the general practitioner for chronic non-specific back and neck complaints, (in this volume)

Mameren H van (1988), Motion patterns in the cervical spine, Thesis, University of Limburg, Maastricht, The Netherlands

Meinert CL (1986), Clinical trials: design, conduct and analysis, Oxford University Press, New York