

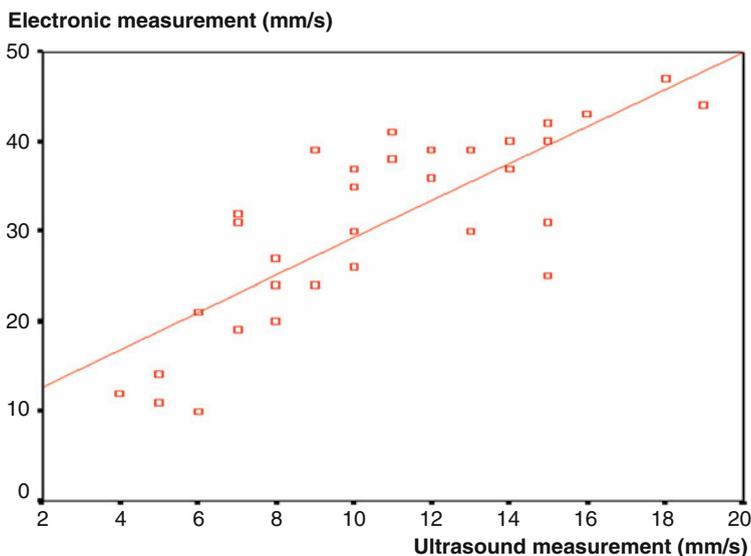


### 3 Primary Scientific Question

Is a new diagnostic test adequately accurate and adequately reliable.

### 4 Testing Accuracy with $R^2$ -Values

Linear regression is often used for that purpose. The underneath figure gives an example. The regression equation is given by  $y = a + b x$  ( $a$  = intercept,  $b$  = regression coefficient). More information of linear regression is given in the Chap. 8. In the underneath example given, the  $x$ -axis-data, ultrasound estimates, are a very significant predictor of the  $y$ -axis-data, the electromagnetic measurements. However, the prediction, despite the high level of statistical significance, is very imprecise. E.g., if  $x = 6$ , then  $y$  may be 10 or 21, and, if  $x = 7$ , then  $y$  may be 19, 31 or 32.



Probably, the best accuracy assessment is to test whether the 95 % confidence interval of the  $a$ -value (the intercept) is significantly different from 0, and whether the 95 % confidence interval of the  $b$ -value (the regression coefficient) is significantly different from 1. However, second best and more easy on a pocket calculator is to test the data for a squared correlation coefficient ( $r^2$ )  $> 95$  %. This method assumes a diagnostic test with the best fit equation  $y = a + b x$ , rather than  $y = x$ . A diagnostic test with the former best fit equation, like in the above example, is not necessarily useless, and could be approved as a valid test, if it is precise, that means,

if the x-data precisely predict the  $(y-a)/b$ - data rather than the y-data. If we apply such a test, then the result of the x-data will, of course, have to be transformed into  $a + b x$  in order to find the y-data.

We will use Electronic Calculator (see Chap. 1) for computations. First, we will calculate the b and r values.

Command:

click ON....click MODE....press 3....press 1....press SHIFT, MODE, and again 1....  
 press = ....start entering the data. ... [x-datum<sub>1</sub> y-datum<sub>1</sub>]. ...[x-datum<sub>2</sub>  
 y-datum<sub>2</sub>]. ...[.....] etc...

In order to obtain the r value, press: shift, S-VAR, ►, ►, 3, = .

The r-value equals 0.6 . . .

The r<sup>2</sup> -value equals 0.36 . . .

This r<sup>2</sup>-value is much smaller than 0.95. It means that this diagnostic test can not be validated as being adequately accurate.

## 5 Testing Reliability with Duplicate Standard Deviations

The reliability, otherwise called reproducibility, of diagnostic tests is another important quality criterion. A diagnostic test is very unreliable, if it is not well-reproducible. Underneath a first data example is given.

test 1	test 2	difference	(difference) <sup>2</sup>
1	11	-10	100
10	0	10	100
2	11	-9	81
12	2	10	100
11	1	10	100
1	12	-11	121
mean			
6.17	6.17	0	100.3

$$\begin{aligned}
 \text{Duplicate standard deviation} &= \sqrt{(1/2 \times \text{mean of (difference)}^2)} \\
 &= \sqrt{(1/2 \times 100.3)} \\
 &= 7.08
 \end{aligned}$$

The proportional duplicate standard deviation

$$\begin{aligned}
 &= \frac{\text{duplicate standard deviation}}{\text{overall mean}} \times 100 \% \\
 &= \frac{7.08}{6.17} \times 100 \% \\
 &= 115 \%
 \end{aligned}$$

An adequate reliability is obtained with a proportional duplicate standard deviation of 10–20 %. In the current example, although the mean difference between the two tests equals zero, there is, thus, a very poor reproducibility.

Underneath a second example is given. The question is, is this test well reproducible?

test 1	test 2
result	
6.2	5.1
7.0	7.8
8.1	3.9
7.5	5.5
6.5	6.6

Analysis:

Test 1	Test 2	Difference	Difference <sup>2</sup>
Result			
6.2	5.1	1.1	1.21
7.0	7.8	−0.8	0.64
8.1	3.9	4.2	17.64
7.5	5.5	2.0	4.0
6.5	6.6	−0.1	0.01
Mean			
7.06	5.78		4.7
grand mean	6.42		

$$\begin{aligned} \text{Duplicate standard deviation} &= \sqrt{(\frac{1}{2} \times 4.7)} \\ &= 1.553 \end{aligned}$$

$$\begin{aligned} \text{Proportional duplicate standard deviation \%} &= \frac{\text{duplicate standard deviation} \times 100 \%}{\text{overall mean}} \\ &= \frac{1.553}{6.42} \times 100 \% \\ &= 24 \% \end{aligned}$$

A good reproducibility is between 10 and 20 %. In the above example reproducibility is, thus, almost good.

## 6 Conclusion

In the current chapter two methods for validation of diagnostic methods with continuous data easily performed with the help of a pocket calculator are described. Many more methods exist. For accuracy assessments paired t-tests, Bland-Altman

plots and the complex linear regression models of Passing-Bablok and Deming are available. For reliability repeatability coefficients and intraclass correlations are possible (see underneath “Note” section). These method is generally more laborious, particularly, with large samples, but available through S-plus, Analyse-it, EP Evaluator, and MedCalc and other software programs.

## **7 Note**

More background, theoretical and mathematical information of validity assessments of diagnostic tests with continuous outcomes is given in *Statistics applied to clinical studies* 5th edition, the Chaps. 45 and 50, Springer Heidelberg Germany, 2012, from the same authors.