Statistical Methods for Comparison of Results from Alternative Methods

Patricia Parreiras
National Institute of Quality Control in Health
FIOCRUZ

March/2005

Some Definitions of Statistics

- Science of collecting and representing data.

- Science that deals with collection of data on a relatively small scale to form logical conclusions about the general case.

- Science of decision-making in the face of uncertainty.

- Science and art of treating data.

Review of Concepts - 1

Linear regression

- Simple linear regression involves discovering the equation for a line that most nearly fits the given data.

-The linear equation is then used to predict values from the data.

It is a mathematic relationship between two or more variables.

Review of Concepts - 2

Correlation

- **Describes the strength, or degree, of linear relationship.**

- **Lets us specify *to what extent the two variables behave alike or vary together.***

Review of Concepts - 3

Correlation coefficient (*r*)

-**Measures the strength and the direction of a linear relationship between two variables.**

Review of Concepts - 4

Coefficient of determination (r^2)

- Represents in what percent the linear model explains the variability of the dependent variable *y*.

- Example,

if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in *y* can be explained by the linear relationship between *x* and *y* (as described by the regression equation). The other 15% of the total variation in *y* remains unexplained.

Limitations of *r* in Methods Comparison

- **Examination of published papers (Dewitte, 2002) showed that most authors were using *r* for this purpose.**

- **Some authors understand that this method does not assess agreement, but association, and that a high *r* does not guarantee good agreement between results.**

Developments

- **Increasing use of an alternative to *r* has been recently detected in literature:**

  - **From 8% in 1995 to 14% in 1996, and to 31-36% in more recent years.**

Altman and Bland (1986) - 1

In clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they agree sufficiently for the new to replace the old.

Some analyses are inappropriate, notably the use of r.

Altman and Bland (1986) - 2

When two methods are compared, we need to assess the degree of agreement. But how?

Altman and Bland (1986) - 3

Many studies give *r* between the results of the two measurement methods as an indicator of agreement.

Most of the analysis were illustrated in this paper by a set of data collected to compare two methods of measuring peak expiratory flow rate (PEFR).

Peak Expiratory Flow Rate (PEFR)

This is a simple method of measuring airway obstruction and it will detect moderate or severe lung disease. It is measured using a standard Wright Peak Flow Meter or a mini Wright Meter.

Altman and Bland (1986) - 4

First step: Plotting data

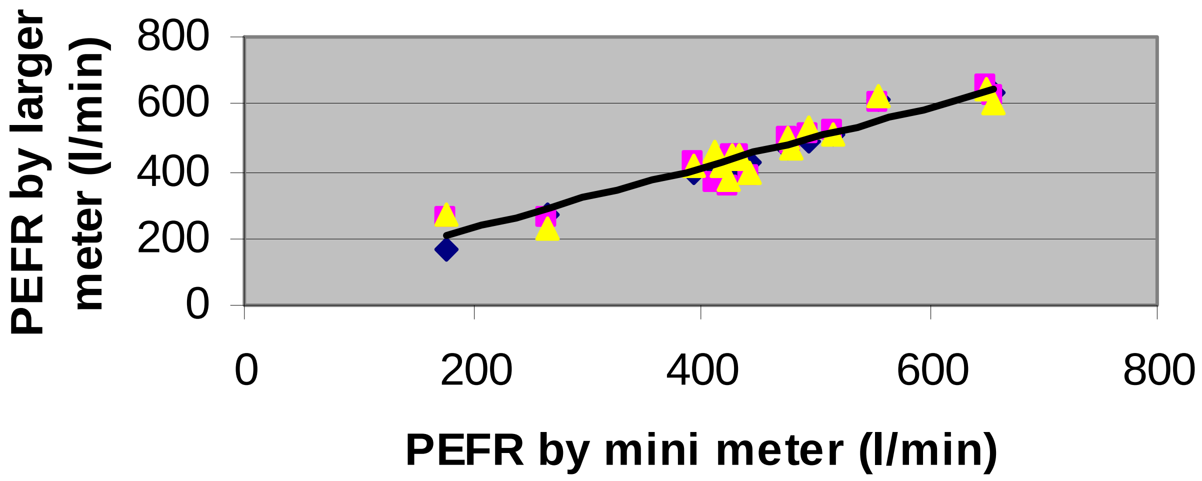

Fig 1: PEFR measured with large Wright peak flow meter and mini Wright peak flow meter, with line of equality.

Altman and Bland (1986) - 5

Second step: **Calculate *r* between the two methods.**

We can safely conclude that PEFR measurements by the mini and large meters are related.

However, a high correlation (0.94 in this example) does not mean that the two methods agree.

Altman and Bland (1986) - 6

Inappropriate use of correlation coefficient

r measures the strength of a relation between two variables, not the agreement between them.

We would have perfect agreement only if the points in Fig 1 lie along the line of equality (slope = 1; intercept = 0).

However, we will have perfect correlation if the points lie along any straight line (any slope; any intercept).

Altman and Bland (1986) - 7

Measuring agreement

It is most unlikely that different methods will agree exactly, by giving identical result for all individuals.

We want to know by how much the new method is likely to differ from the old:

> If this is not enough to cause problems in clinical interpretation, we can replace the old method by the new or use the two interchangeably.

Altman and Bland (1986) - 8

If the two PEFR meters were unlikely to give readings which differed by more than, say, 10 l/min, we could replace the large meter by the mini meter because so small a difference would not affect decisions on patient management.

Altman and Bland (1986) - 9

On the other hand, if the meters could differ by 100 l/min, the mini meter would be unlikely to be satisfactory.

How far apart measurements can be without causing difficulties will be a question of judgment.

Ideally, it should be defined in advance to help in the interpretation of the methods comparison.

Altman and Bland (1986) -10

Then….

The first step is to examine the data.

A simple plot of the results of one method against those of the other (Fig 1) though without a regression line (scatter plot) is a useful start.

Nevertheless usually the data points will be clustered near the line and it will be difficult to assess between-method differences.

Altman and Bland (1986) - 11

A plot of the difference between the methods against their mean may be more informative.

Altman and Bland (1986) - 12

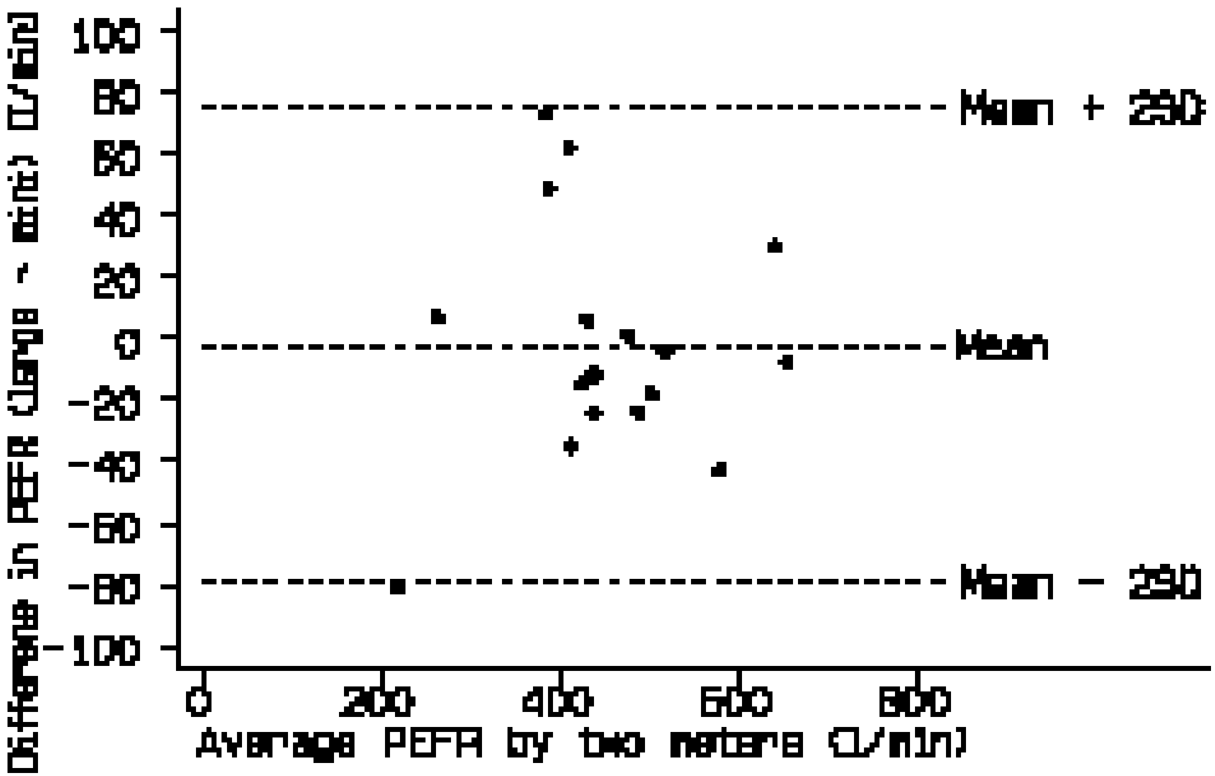

Fig 2. Difference against mean for PEFR data.

Altman and Bland (1986) - 13

Fig 2 displays considerable lack of agreement between the large and mini meters, with discrepancies of up to 80 l/min; these differences are not obvious from Fig 1.

Altman and Bland (1986) - 14

In the analysis of measurement method comparison data, neither the correlation coefficient (as we show here) nor techniques such as regression analysis are appropriate.

The paper suggest replacing these misleading analyses by a method that is simple both to do and to interpret.

Altman and Bland (1986) - 15

Why has a totally inappropriate method, the correlation coefficient, become almost universally used for this purpose?

Two processes may be at work here - namely, pattern recognition and imitation.

Altman and Bland (1986) - 16

Once the correlation approach has been published, others will read of a statistical problem similar to their own being solved in this way and will use the same technique with their own data.

Medical statisticians who ask:

“Why did you use this statistical method?"

Will often be told:

“Because this published paper used it".

PATRICIA PARRREIRAS

Research Mentor: Juan Arciniega
LMDCQ/CBER/FDA

FUNDACAO OSWALDO CRUZ - BRASIL

National Institute of Quality Control in Health - BRASIL

Belo Horizonte – Minas Gerais State

THANK YOU VERY MUCH!!!