

**Friday Seminar 3/10/06**

**Measures of Reproducibility – Part 1**

**Steven Lewis, MS, MBA**

This discussion will focus on describing methods used for comparing measurements obtained from two different instruments, time points, or raters. We will begin by briefly discussing familiar measures of association and agreement such as Pearson's Correlation Coefficient (R) and Cohen's Kappa ( $\kappa$ ).

The core of this first talk focuses on graphical (Bland-Altman, mountain plots) as well as model-based methods (Passing-Bablok models, Deming regression) for detecting measurement bias. All of these methods are implemented in MedCalc, a method comparison software package licensed by the Center for Health Care Research and Policy.

Pros and cons of the various methodologies for assessing a measurement's reproducibility will be identified through two real-data examples. In the first, we examine the problem of assessing the reliability of laboratory hemoglobin A1c measurements obtained for a series of samples using two different machines. In the second, in collaboration with Tyler Stevens and colleagues at the Cleveland Clinic, we employ several reliability assessments to help determine optimal times to collect duodenal aspirates in the development of a shortened pancreatic function test.

The second part of this presentation (coming in a few weeks), will provide a more in-depth analysis of Kappa Coefficients, and describe alternative methods for measuring reproducibility, specifically intraclass correlation coefficients (ICC) and approaches based on structural equation modeling (SEM).

This work is a group effort, including contributions from Charles L. Thomas, Adam T. Perzynski, Thomas E. Love, and Neal V. Dawson.

Steven Lewis [sal25@case.edu]