1

Introduction

The classical approach to analysis of method comparison studies is the Bland–Altman plot where differences between methods are plotted against averages, leading to the limits of agreement and to verification of whether the underlying assumptions are fulfilled. This plot is merely a 45° rotation of a plot of the methods versus each other, while the limits of agreement correspond to prediction limits for the conversion between the methods.

This one-to-one correspondence between a prediction interval for the difference between two methods and the prediction of a measurement by one method given a measurement by the other is in this book carried over to an explicit modeling of data with the aim of producing conversion equations between methods.

The explicit definition of a model generating the data obtained is virtually absent in the literature. The aim of this book is to fill this gap. By explicitly defining a model for the data it is possible to discuss relevant quantities to report and their interpretation and underlying assumptions, without involving technicalities about estimation.

It is my opinion that presentation of concepts in terms of a statistical model enhances understanding, because it allows the technicalities about estimation procedures to be relegated to technical sections, and thereby allows the interpretation of models and the correspondence with practice to be discussed free of technicalities. Conversely, it is also possible to discuss estimation problems more precisely when a

Comparing Clinical Measurement Methods: A Practical Guide Bendix Carstensen

well-defined model is specified. An explicitly defined model also makes it possible to simulate data for testing proposed measures and procedures.

The purpose of introducing explicit models is therefore not to give a formalistic derivation of all procedures, but rather to give a framework that can be used to assess the clinical relevance of the procedures proposed.

The technical sections of this book assume that the reader is familiar with standard statistical theory and practice of linear models as well as of random effects (mixed) models. However, a lack of skills should not be a major impediment to understanding the general ideas and concepts.

The core assumption in the models used in this book is that conclusions concerning the methods compared should not depend on the particular sample used for the comparison study. Taken to the extreme this is of course never true, but my point is that the particular distribution of blood glucose, say, among patients in a study should not influence conclusions regarding relationships between different methods to measure it. Samples chosen for method comparison studies should reflect the likely *range* in which comparisons are used in the future. Any attempt to make the sample used for the method comparison study representative of future *distribution* in samples where the results are applied is futile and irrelevant.

In statistical terms this means that models presented all have a systematic effect of item (individual, sample). Moreover, this point of view automatically dismisses all measures based on correlation. Hence, such measures are only mentioned briefly in this book.

The aim of the book is to give the reader access to practical tools for analyzing method comparison studies, guidance on what to report, and perhaps most importantly some guidance on how to plan comparison studies and (in the event this is not followed) hints as to what can and what cannot be inferred from such studies, and under what assumptions. An extensive treatise on general measurement problems can be found in Dunn's book [15].

The book starts with a few brief examples that highlight some of the topics in the book: (1) the simplest situation, with one measurement by each of two methods; (2) replicate measurement by each method and exchangeability; (3) linear relationship with slope different from 1; and (4) more than two methods.

The next chapter treats the situation with one measurement per individual by two methods in more depth, mentioning some of the more common methods of regression with errors in both variables. Chapter 5 treats the case where replicate measurements are taken on each individual, and gives advice on how to treat the situation with standard software.

The core of the book is Chapter 7, with the exposition of a general model that contains all the previous models as special cases. The model is expanded using transformation of data in Chapter 8.

What is *not* treated in this book are models for completely general non-linear relationships between measurement methods, except in so far as they can be transformed to the linear case. Likewise, the case of non-constant variances is also only treated in cases where data can be transformed to the constant variance case.

All graphs in this book are generated by R, and most are the result of functions specially designed to handle method comparison data collected in the package `MethComp` developed by Lyle Gurrin and me. The majority of the procedures in Chapters 4 and 5 can fairly easily be implemented in existing standard software. Examples of code for these methods are given in Chapter 12 for SAS, Stata and R.

When non-constant bias is introduced the underlying models become largely intractable, and the only viable method of estimation in finite (programming) time is to use either the ad-hoc procedure of alternating regressions or the `BUGS` machinery in one of the available implementations. The models proposed are wrapped up in the `MethComp` package for R.

There is a website `http://www.biostat.ku.dk/~bxc/MethComp` for the `MethComp` package where examples and illustrative programs can be found. The website also contains links to teaching material related to this book, including practical exercises with corresponding solutions.