Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 3383–3414
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Assessing a vector of clinical observations

Morris L. Eaton[a], Robb J. Muirhead[b,*], Eve H. Pickering[c]

[a] *School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455, USA*
[b] *Statistical Research and Consulting Center, Pfizer Inc., MS6025-A5115, 50 Pequot Ave, New London, CT 06320, USA*
[c] *Nonclinical Statistics, Pfizer, Inc., MS8260-2210, Eastern Point Road, Groton, CT 06340, USA*

Received 16 February 2004; received in revised form 23 December 2004; accepted 8 February 2005
Available online 25 April 2005

Abstract

The periodic monitoring of drug treatments often involves the collection of biological specimens (e.g. blood, urine, synovial fluid) for the purpose of clinical laboratory assessment. The analysis of a particular specimen yields a vector of measurements from which judgments are made concerning the status of a subject and the effect of the drug. Typically, an observation vector is compared to "normal values" which may be conditioned on covariates such as age, gender, or other relevant characteristics. Under an assumption of multivariate normality of the data available, a method is presented for deciding whether a particular observed vector looks "normal". The method, based on a predictive approach, is compared to other proposals and is shown to have optimality properties not possessed by standard procedures. Three different approaches are used in the discussion of optimality within the class of invariant methods. The first involves tolerance regions with smallest normalized expected volume, the second involves a decision theoretic comparison of predictive distributions, while the third involves the foundational notions of incoherence (Dutch book) and strong inconsistency.

Keywords: Clinical observation vectors; Prediction; Tolerance regions; Optimality; Invariance; Decision theory; Incoherence (Dutch book)

* Corresponding author.
E-mail address: robb.j.muirhead@pfizer.com (R.J. Muirhead).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.02.019

1. Introduction

The primary motivation for the methodology introduced in this paper is well illustrated by a standard clinical trial protocol. Almost every clinical trial includes the collection of biological specimens (e.g. blood, urine, synovial fluid) for the purpose of clinical assessment. The usual battery of laboratory tests often involve many separate assessments, broken into panels of similar tests, such as blood chemistry, liver function and kidney function tests. The results from these laboratory tests are used to assess safety and efficacy of a drug in a clinical trial setting. Laboratory values for a subject are compared to a "normal" range, claimed by each laboratory to be obtained from data from a "healthy normal" population. These data may be specific to age, gender and other characteristics thought to be relevant. The comparison typically results in either the subject being classified as "within the normal range" or "outside the normal range". This latter classification may result in some intervention.

Certain patterns of laboratory abnormalities may indicate particular toxic effects while others identify the disease state and still others may be benign. Some patterns may indicate the efficacy of a drug. A particularly illuminating example concerns the use of liver panel data to assess the possibility of liver toxicity (hepatatoxicity) in a subject. The panel measurements would ordinarily include laboratory results for aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin, alkaline phosphatase (ALP), and lactate dehydrogenase (LDH) among others. Since ALT is considered to be the most specific for drug-induced liver toxicity, this would likely be the first result looked at by an investigator. If ALT were elevated, the investigator might next consider AST, which is highly correlated with ALT. This results in a conditional (given the value of ALT) assessment of the AST value. If AST is not also elevated, the clinician may suspect laboratory error and request a repeat measurement. If both ALT and AST are elevated, the clinician might next consider total bilirubin, followed by ALP. As more tests on the list are elevated, the indication of drug-related liver toxicity becomes stronger. Similar considerations apply to the evaluation of a vector of observations made to assess renal toxicity.

Our approach to the problem of assessing clinical observations is based on methodology obtained from ideas in statistical prediction. It is assumed that the data D, known to originate from a population of "normal" individuals, are available. Using this data, methods of statistical prediction theory can often be used to construct either prediction regions and/or predictive distributions for a contemplated new observation Z from the population of "normals". An observed value $Z = z_0$ can now be assessed as "normal" or "not normal" by checking the value z_0 against what was predicted by the prediction regions and/or the predictive density.

Throughout this paper attention is restricted primarily to the case where the data D are a random sample $X_1, \ldots, X_n$ from a p-dimensional normal distribution, $N_k(\mu, \Sigma)$, with an unknown mean vector μ and an unknown $k \times k$ positive definite covariance matrix Σ. In this case, a predictive region for the next observation $Z = X_{n+1}$ was proposed by Fraser and Guttman (1956) and called a tolerance region by them. To describe this region, let $X = (X_1, \ldots, X_n)$ be the data, let $\bar{X}$ be the sample mean, and let

$$\tilde{S} = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'$$