Chemometrics and Intelligent Laboratory Systems 99 (2009) 49–56

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Using the similarity factor f_2 in practice: A critical revision and suggestions for its standard error estimation

Jordi Ocaña [a,*], Gloria Frutos [b], Pilar Sánchez O [c]

[a] Department of Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain
[b] Department of Statistics and Operational Research, Faculty of Pharmacy, Complutense University of Madrid, Madrid, Spain
[c] Faculty of Sciences Chemicals and Pharmaceuticals, University of Chile, Santiago de Chile, Chile

ARTICLE INFO

Article history:
Received 30 March 2009
Received in revised form 23 June 2009
Accepted 19 July 2009
Available online 29 July 2009

Keywords:
Simulation
Bias correction
Bootstrap
Jackknife
Consumer risk
Producer risk

ABSTRACT

The purpose of this research was to develop new procedures with the aim of improving the usage of the similarity factor f_2 in dissolution data analysis, and to evaluate them jointly with preexisting ones. We introduce bias-correction and standard error estimation procedures based on the delta, the jackknife and the bootstrap methods. These methods, jointly with the rule of declaring similarity when f_2 exceeds 50 and some alternative testing procedures based on bootstrap confidence intervals, are evaluated on experimental data and studied by simulation. The results indicate that no method is strictly the best, but the following conclusions seem to appear: for estimation purposes the most reliable approach is to use the plain sample f_2 instead of any bias-corrected alternative, any of the standard error estimates may be used in practice and, most importantly, there are evidences against the validity of the procedure declaring similarity if the sample f_2 exceeds 50; a decision rule based on a confidence interval seems to be more adequate. In any case the question should be further investigated.

1. Introduction

A dissolution assay is a part of the demonstration of dosage form ability to deliver in vivo the drug substance, so is the most important test to show the biopharmaceutical quality. Also dissolution assays can serve other purposes: during the development of a medicinal product a dissolution test may be used as a tool to identify formulation factors that are influencing the bioavailability of the drug; in the quality control of scale-up and of production batches, a dissolution test may be used to prove consistency in the manufacturing and to ensure that the dissolution profiles remain similar to those of pivotal clinical trial batches. Furthermore, in bioequivalence surrogate inference a dissolution test can be used to demonstrate similarity between different formulations of an active substance and the reference medicinal product. Consequently, interpretation and understanding of dissolution results is an area of interest by the industry and the regulatory authorities as well as by the pharmaceutical research.

According to Tsong et al. [1], dissolution data treatment may be classified in model-dependent methods, model-independent methods and statistical procedures such as multivariate analysis approach, method of repeated measures, time series approach and others. The suitability of the mathematical model is determined by non-linear regression analysis; once a suitable function has been selected, the dissolution behavior is evaluated according to the physical meaning of the fitted parameters. Model independent methods include the difference factor f_1 and the similarity factor f_2, [2] that compare the dissolution profiles of a pair of drug products using the dissolution data in their native form.

It is evident from the pharmaceutical literature that no single approach is widely accepted to determine if dissolution profiles are similar [3]. Nevertheless, the similarity factor f_2 (a descriptive index) is gaining popularity due to its recommendation by various regulatory committees as a criterion for the assessment of the similarity between two dissolution profiles.

The U.S. Food and Drug Administration (FDA) has adopted the similarity factor as a simple method to compare dissolution profiles, in order to avoid subjective evaluation of dissolution profile comparison. SUPAC-IR (1995) Guidance for Industry, [4], was the first regulatory document from FDA to establish a clear policy for several scale-up and post approval changes, applied to immediate release solid oral dosage forms. According to this guidance, dissolution profiles may be compared using the similarity factor f_2 and all profiles should be conducted on at least 12 individual dosage forms. Guidance [5] is an extension of SUPAC-IR providing criteria for qualifying biowaivers for drug products applied to active pharmaceutical ingredients exhibiting high solubility and high permeability.

In the European Community, in addition to European Pharmacopoeia, there are several guidelines on dissolution assays. The European Medicines Agency suggests using the similarity factor f_2 to compare dissolution profiles in Refs. [6] and [7].

* Corresponding author. Departament d'Estadística, Facultat de Biologia, Av. Diagonal, 645, 08028 Barcelona, Spain. Tel.: +34 934021565; fax: +34 934111733.
E-mail address: jocana@ub.edu (J. Ocaña).

0169-7439/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2009.07.010

The above regulatory suggestions, possibly in addition to its simplicity, explain why f_2 is widely used in practice. In applications, f_2 is mainly used as a:

a) Response or dependent variable [8–16] usually with optimization purposes, e.g. to compare manufacturing processes for establishing experimental conditions maximizing similarity between formulations.
b) Part of a decision criterion to establish similarity of two formulations [17,18]. The regulatory suggestion "decide similarity if (the sample) f_2 exceeds 50" is applied in a literal sense.

Seldom is the sample f_2 value complemented with an indication of its precision, like a confidence interval, and sometimes its origin is unclear [19] when it is included.

In the "a)" case, bias with different sign under two experimental conditions would artificially enlarge or diminish the differences in the observed response. The use of f_2 as a response variable is critically considered in [20]. In the "b)" case, the type I and II error probabilities (that is, the user and producer risks, respectively) may not be adequately under control. If in fact the *true* f_2 is less than 50, the probability of observing from experimental data a *sample* f_2 exceeding 50 is not zero, and should be calculated. The need for more statistically sound approach seems evident.

The next section, "methods", is divided in two subsections: the first one summarizes the main concepts on statistical inference based on f_2, and introduces some additional methods; the second subsection describes our simulation approach to study the properties of all these statistical methods. The "results" section presents some computations on real datasets and the simulation results. The "discussion" and "conclusion" sections try to highlight the main consequences on the similarity factor applicability in practice. In the supplementary material web page http://www.ub.es/stat/angles/recerca/materials/f2_supplementary.htm there are the complete datasets and the mathematical details in the genesis of the new statistical indexes.

2. Methods

2.1. Statistical inference based on the similarity factor

Typically, dissolution data are repeated measures where drug percent dissolved is the dependent variable and time is the repeated factor. More formally, the response variable in dissolutions assays, x_{ijk}, is the cumulative percentage dissolved from tablet or unit k ($k=1,\ldots,n$), at sampling time t_j ($j=1,\ldots,P$) of the test or reference batch, ($i=T,R$). These observations may be collected in a three-way array, $\mathbf{x}=(x_{ijk})_{i=T,R;\ j=1,\ldots,P;\ k=1,\ldots,n}$.

The (sample) similarity factor is defined as:

$$\hat{f}_2 = 50\log\left\{\left[1+\frac{1}{P}\sum_{j=1}^{P}(R_j-T_j)^2\right]^{-1/2}\cdot 100\right\} \quad (1)$$

where $R_j=(\Sigma_{k=1}^{n}x_{Rjk}/n)$ and $T_j=(\Sigma_{k=1}^{n}x_{Tjk}/n)$ correspond to the sample means for all tablets at time point j, for the reference and test batches, respectively. Here "log" stands for the base 10 logarithm.

The sample f_2 and all indexes considered here are functions of the average differences $\hat{\boldsymbol{\delta}}=(\hat{\delta}_1,\ldots,\hat{\delta}_p)'$ with $\hat{\delta}_j=R_j-T_j$. For brevity, we will express all these indexes in terms of $\boldsymbol{\delta}$ and of

$$d(\hat{\delta}) = 1+\frac{1}{P}\sum_{j=1}^{P}\hat{\delta}_j^2. \quad (2)$$

For example, the sample f_2 becomes $\hat{f}_2 = 50\log\left\{[d(\hat{\delta})]^{-1/2}\cdot 100\right\} = 100-25\log\{d(\hat{\delta})\}$. Table 1 summarizes all the f_2 indexes considered in this paper.

Note that the variance of the $\hat{\delta}_j$ differences can be estimated by:

$$\widehat{\mathrm{var}}(\hat{\delta}_j) = (s_{Rj}^2+s_{Tj}^2)/n \quad (3)$$

where s_{Rj}^2 and s_{Tj}^2 are the unbiased estimates of the variance for all data at the *j-th* time point, for the reference and test batches, respectively.

To what extent conclusions based on the $\hat{f}_2$, are reliable? Its direct use normally ignores that it is just an estimate, subject to statistical error, of the *true* f_2 parameter:

$$f_2 = 50\log\left\{\left[1+\frac{1}{P}\sum_{j=1}^{P}(\mu_{Tj}-\mu_{Rj})^2\right]^{-1/2}\cdot 100\right\} \quad (4)$$

where μ_{Tj} and μ_{Rj} correspond to the *true* "population" means.

Liu et al. [21] discuss the methodological problems associated to the use of the *sample* f_2 statistic, $\hat{f}_2$. Provided that the only experimentally observable quantity is $\hat{f}_2$, both to state the alternative hypothesis of similarity as H_1: "$\hat{f}_2>50$" or to use the decision rule "reject the null hypothesis of no similarity if $\hat{f}_2>50$" are ill-defined. These authors establish that $\hat{f}_2$ is a biased but asymptotically unbiased and consistent estimator of f_2. They suggest the index $\hat{f}_2^*$ (see also Table 1) as a bias-corrected version of $\hat{f}_2$.

Provided that it is very difficult to obtain the sampling distribution of $\hat{f}_2$, or even some of its characteristics like its standard error, to test for dissolution profile similarity using f_2, Shah et al. [22] use nonparametric "bootstrap bias corrected and accelerated", BCa, confidence intervals.

Ma et al. [23] state in a precise way the problem of testing for similarity of dissolution profiles using the similarity factor. These authors distinguish between the "population" or "theoretical" f_2 value, specifying hypotheses about similarity, and the sample f_2, $\hat{f}_2$,(or related statistics) to be used in estimating the true f_2 or in testing hypotheses about it. Under this view, deciding about similarity between dissolution profiles may be interpreted in terms of a null hypothesis of dissimilarity vs. an alternative hypothesis of similarity:

$$H_0: f_2\le\theta_0 \text{ vs. } H_1: f_2>\theta_0 \quad (5)$$

where θ_0 is a similarity limit, typically $\theta_0=50$. This is also the point of view taken in our paper.

Ma et al. use the simpler "bootstrap percentile" confidence intervals. The associated decision rule is "decide similarity if $L_P^\alpha>\theta_0(=50)$" where L_P^α stands for the lower confidence limit of the one-sided percentile interval with nominal confidence$(1-\alpha)100\%$, $[L_P^\alpha, 100]$. Their simulation results indicate that $\hat{f}_2$ is not very biased and in general better than $\hat{f}_2^*$, which overcorrects for bias. The test based on percentile bootstrap intervals seems to perform adequately in terms of type I error probability and acceptably in terms of type II error probability. The simulated dissolution profiles are generated according to a normal distribution with independent observations inside each simulated dosage unit. One of the objectives of our paper is to elucidate if these results are still valid when some dependence structure between the observations inside each dosage unit is introduced, in agreement with what seems to happen in real dissolution experiments. The other objective is to introduce some additional indexes related to f_2 and to determine if these additional indexes represent a possible improvement.

A possible drawback of the bias-corrected estimator $\hat{f}_2^*$ is that it is undefined when $\sum_{j=1}^{P}(R_j-T_j)^2<\sum_{j=1}^{P}(s_{Rj}^2+s_{Tj}^2)/n$. This is overcome by the alternative bias-corrected estimator $\hat{f}_{2,bc}$ introduced in Table 1. This bias-corrected estimator tries to take into account the within-units correlation. Under complete profile equality, $\hat{\boldsymbol{\delta}}=\mathbf{0}$ the last term of $\hat{f}_{2,bc}$ vanishes and it is always less than 100. This counterintuitive trend is palliated by growing values of n and/or P.

An alternative approach for bias correction, slightly more computer intensive than the preceding one, may be based on the jackknife method. The resulting estimator $\hat{f}_{2,Jbc}$ is also shown in Table 1. According to the usual jackknife procedure and notation, what is designated $\hat{f}_{2,(R-j)}$ in Table 1 corresponds to the standard estimate Eq. (1) but computed deleting the j-th unit or tablet profile in the R data, and similarly for $\hat{f}_{2,(T-j)}$. A counterintuitive aspect of $\hat{f}_{2,Jbc}$ is that