Journal of Pharmaceutical and Biomedical Analysis 88 (2014) 27–35

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

A comparison of methods for classifying samples as truly specific with confirmatory immunoassays

Thomas Jaki [a,*], John-Philip Lawo [b], Martin J. Wolfsegger [c], Peter Allacher [c], Frank Horling [c]

[a] Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom
[b] CSL Behring GmbH, Emil-von-Behring-Straße 76, 35041 Marburg, Germany
[c] Baxter Innovations GmbH, Wagramer Strasse 17-19, 1220 Vienna, Austria

ARTICLE INFO

Article history:
Received 13 March 2013
Received in revised form 5 August 2013
Accepted 6 August 2013
Available online 21 August 2013

Keywords:
Anti-drug antibody
Confirmatory
Cut point
Immunoassay
Specificity

ABSTRACT

Biotechnology-derived therapeutics may induce an unwanted immune response leading to the formation of anti-drug antibodies (ADAs) which can result in altered efficacy and safety of the therapeutic protein. Anti-drug antibodies may, for example, affect pharmacokinetics of the therapeutic protein or induce autoimmunity. It is therefore crucial to have assays available for the detection and characterization of ADAs. Commonly, a screening assay is initially used to classify samples as either ADA positive or negative. A confirmatory assay, typically based on antigen competition, is subsequently employed to separate false positive samples from truly positive samples. In this manuscript we investigate the performance of different statistical methods classifying samples in competition assays through simulation and analysis of real data. In our evaluations we do not find a uniformly best method although a simple *t*-test does provide good results throughout. More crucially we find that very large differences between uninhibited and inhibited measurements relative to the assay variability are required in order to obtain useful classification results questioning the usefulness of competition assays with high variability.

1. Introduction

The development of anti-drug antibodies against biotechnology derived therapeutics displays a major obstacle in drug development. Therefore, validated assays and strategies for the detection and characterization of anti-drug antibodies (ADAs) against a protein therapeutic are required to monitor the unwanted immune response. Several regulatory guidelines [1–3] and white papers [4–6] have been published describing testing strategies, assay formats, validation requirements and performance expectations for such assays. The overall strategy comprises of an initial screening assay followed by a confirmation assay. The screening assay allows distinguishing between negative and – true or false – positive samples, while the subsequent confirmatory assay will differentiate between true and false positive samples.

A critical step during assay development is the establishment of cut points that are used to classify samples into negative and positive samples. Using a risk-based approach, an upper negative limit of 95% for the screening cut point is recommended [4,6], resulting in a 5% false-positive rate. Subsequently, the confirmation assay aims to eliminate false positive samples. Recently various methods for finding cut points for screening assays have been evaluated [7,8], while no such work to date exists for classifying confirmatory samples.

The confirmation assay is usually a competitive inhibition assay evaluating the signal differences of a sample with or without preincubation of the therapeutic drug. A sample is considered as positive when a decrease in assay signal in the presence of therapeutic drug is observed. The principle of this approach is widely accepted, while determining how much the assay signal needs to be reduced to evaluate a sample as positive has to be established for each single assay. The confirmation cut point may be established based on the specific signal reduction of a positive sample in the confirmatory assay or based on the background variation of untreated healthy population or patients.

General strategies and recommendations to establish a confirmation cut point were described in [9,10]. However, the need for publications exploring the implementation of appropriate and valid methodology was recommended by Smith et al. [9]. In this paper, we describe and evaluate different scenarios and statistical approaches to classify data from competition assays and discuss weakness and strength of each approach.

2. Classifying samples

In this section we will describe four different methods for classifying samples in competition assays. Shankar et al. [6] describe

* Corresponding author. Tel.: +44 (0)1524 59 23 18.
E-mail address: jaki.thomas@gmail.com (T. Jaki).

0731-7085/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jpba.2013.08.013

the first three methods, namely the parametric difference method, the fixed percent inhibition cut point and the parametric % inhibition methods, although only the latter is recommended in that paper. The fourth approach, a simple t-test, is introduced by Neyer et al. [10]. Notice that the parametric difference and the parametric % inhibition method use drug-naïve samples to establish a cut point that is subsequently used to classify samples. The other two approaches are only based on the samples to be classified. In our evaluation of methods a large number of alternative approaches (e.g. prediction intervals [8] and robust parametric method [7]) have also been investigated, but since we found no obvious advantage of any of these methods over the ones described, we will focus our discussion on popular ideas in the literature. The principle idea of each approach is to determine if the change in assay signal with and without pre-incubation of the therapeutic drug is large enough to be a relevant indicator to distinguish between true positive and false positive samples. We will therefore consider the situation where measurements with and without pre-incubation for each sample are available and assume pre-incubation is successful and truly leads to inhibition. Moreover, we assume that multiple runs (analysis) per sample are undertaken and that measurements are corrected for run noise. Measurements with pre-incubation of the therapeutic drug will be referred to as "inhibited measurements" and without incubation as "uninhibited measurements". We begin by describing the basic methods for classification and subsequently discuss the different ideas about using multiple runs.

2.1. Methods for classification

Method 1: parametric difference

For each sample find the difference between uninhibited and inhibited measurement as

$D =$ uninhibited measurement − inhibited measurement.

The cut point, which is typically based on a small number of drug-naïve samples, is found as $c_D = \overline{D} + z_{0.999} * \sigma_D$ where $\overline{D}$ is the average difference across all samples, σ_D is the corresponding standard deviation and $z_{0.999}$ is the 99.9% percentile of the standard normal distribution (approximately 3.09). Note that the 99.9% percentile corresponds to a false-positive rate of 0.1% we are willing to accept [6]. A new, not necessarily drug naïve, sample with observed difference D^* is considered positive if $D^* > c_D$. Notice that an alternative approach log-transforms the measurements prior to taking the difference. Such an approach does perform better if measurements are log-normally distributed, but much worse if they are normal. The presented formulation offers a good balance in both cases as we will see below.

Method 2: parametric % inhibition

For each sample from a small initial pool of drug naïve samples, find the percent change in inhibition as

$$I = 100 * \left(1 - \frac{\text{inhibited measurement}}{\text{uninhibited measurement}}\right)$$

The inhibition based cut point is found as $c_I = \overline{I} + z_{0.999} * \sigma_I$ where $\overline{I}$ is the average percent change in inhibition across all samples, σ_I is the corresponding standard deviation and $z_{0.999}$ is the 99.9% percentile of the standard normal distribution as before. A new, not necessarily drug naïve, sample with inhibition I^* is considered positive if $I^* > c_I$.

Method 3: fixed % inhibition

For each sample find the percent change in inhibition as

$$I = 100 * \left(1 - \frac{\text{inhibited measurement}}{\text{uninhibited measurement}}\right)$$

as for the previous method. A new sample is considered positive if the percent change in inhibition exceeds a fixed values, such as 25 or 50%. In our evaluations we will present results for the latter so that samples are positive if $I > 50$ as this percentage resulted in good classification. Note that, in contrast to the first two methods no initial pool of drug naïve samples is required to classify as no assay specific cut point is used.

Method 4: t-test

For each sample, perform a one-sided 2-sample t-test (see e.g. [11] for details) of all runs of the log-transformed study drug inhibited values against the log-transformed uninhibited values. If the resulting p-value is less than 0.01 the sample is classed positive. Note that the log-transformation is used since typically log-normal distributions are assumed for the assay measurements. We will keep to this recommendation from [10] in our evaluations despite violating this assumption on occasion to explore the impact of the misspecification. Note further that this approach also does not require the initial pool of data to find a cut point but rather calibrates the results for each sample separately.

2.2. Using multiple runs

The final method based on a t-test does require multiple runs per sample to be available, while the first three methods introduced above do not make explicit use of multiple runs in the experiment. For the fixed % inhibition approach one can either classify each run separately or classify the average of the runs only. Should the former approach be used, inconsistencies between results may occur such that samples are classed differently between runs. Combining the information of all runs (e.g. by averaging, fitting a model, etc.) to base classification on is therefore preferred as it avoids these inconsistencies.

For the parametric difference and parametric % inhibition method, one can find an overall cut point either

1. based on an average of the runs per sample (e.g. mean per subject across runs),
2. based on the average of per run cut points (i.e. find a cut point for each run and then average), or
3. based on pooling all data (i.e. runs are treated as independent samples).

The third option is unsuitable for establishing a valid cut point as it makes the assumptions that measurements of the same sample in different runs are independent of each other. This increases the number of available samples used to find the cut point artificially and will consequently result in too small standard deviations, σ_D and σ_I.

The choice between the first two approaches comes down to the subsequent use of the cut point. If samples are to be classified based on a single run, then the average of the per run cut points (option 2) is appropriate while the average per subject across runs (option 1) as the basis for the cut point computation is appropriate if multiple runs are utilized. The reason for this distinction is that typically the variability in a single run is larger than the variability of the average of the runs, which must be taken into account when finding the cut point as well.

For our evaluations below we will use the first option as the results are unambiguous for the fixed % inhibition approach and qualitatively the same as for option 2 for the other methods. In particular we take the mean value of the runs per sample to find the average across subjects (plus the corresponding variance) for simplicity and to maintain focus on the direct comparison of methods. Depending on the design used to obtain the data, more advanced