Journal of Pharmaceutical and Biomedical Analysis 128 (2016) 166–173

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

A false sense of security? Can tiered approach be trusted to accurately classify immunogenicity samples?

Thomas Jaki[a,*], Peter Allacher[b,c], Frank Horling[b]

[a] Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom
[b] Baxalta Innovations GmbH, Industriestrasse 67, 1221 Vienna, Austria
[c] IMC University of Applied Sciences, Piaristengasse 1, 3500 Krems an der Donau, Austria

ARTICLE INFO

Article history:
Received 19 February 2016
Received in revised form 18 May 2016
Accepted 19 May 2016
Available online 27 May 2016

Keywords:
Anti-drug antibody
Confirmatory
Cut point
Immunoassay
Immunogenicity
Screening
Specificity

ABSTRACT

Detecting and characterizing of anti-drug antibodies (ADA) against a protein therapeutic are crucially important to monitor the unwanted immune response. Usually a multi-tiered approach that initially rapidly screens for positive samples that are subsequently confirmed in a separate assay is employed for testing of patient samples for ADA activity. In this manuscript we evaluate the ability of different methods used to classify subject with screening and competition based confirmatory assays. We find that for the overall performance of the multi-stage process the method used for confirmation is most important where a *t*-test is best when differences are moderate to large. Moreover we find that, when differences between positive and negative samples are not sufficiently large, using a competition based confirmation step does yield poor classification of positive samples.

1. Introduction

Detecting and characterizing of anti-drug antibodies (ADA) against a protein therapeutic are crucially important to monitor the unwanted immune response. Usually a multi-tiered approach that initially rapidly screens for positive samples that are subsequently confirmed in a separate assay is employed for testing of patient samples for ADA presence. Several regulatory guidelines [1–3] and white papers [4–6] describe the testing strategies, assay formats, validation requirements and performance expectations for such assays have been published.

In order to use either screening or confirmatory assays, establishing cut points that are used to classify into negative and positive samples are paramount. An upper negative limit of 95% for the screening cut point is recommended [1,2,4,6], resulting in a 5% false-positive rate. The subsequent confirmation assay used here aims to eliminate false positive samples based on competition assays. These competition assays are a tool to identify possible signal contribution from unspecific antibody binding and additionally analyze all samples using a study-drug inhibited assay. This assay is basically set up identically to the uninhibited assay with the exception that all samples are pre-incubated with excess amount of free specific protein antigen ("antigen competition"). Specific antibodies directed against the particular antigen are bound in the form of immune complexes in the liquid phase and subsequently removed during washing steps. Hence, the specificity of antibodies detected with the uninhibited assay can be confirmed by a reduction of signal in the inhibited assay. Recently various methods for finding cut points for screening assays [7,8] and confirmatory assays [9] have been evaluated.

One of the unexpected and striking findings when evaluating the performance of confirmatory assays [9] was that extremely large differences between uninhibited and inhibited samples are necessary to separate positive from negative samples. This surprising finding led us to investigate the capability of the multi-tier approach to separate positive and negative samples. In this manuscript we will evaluate the ability of the multi-tier approach for classifying samples in both simulations and real data evaluations.

2. Classifying samples

Previously a large number of different approaches for classifying screening (e.g. [6,7]) and confirmatory assays (e.g. [9]) have

* Corresponding author. Tel.: +44 01524 59 23 18.
E-mail address: jaki.thomas@gmail.com (T. Jaki).

http://dx.doi.org/10.1016/j.jpba.2016.05.031

been described. In this evaluation we consider 7 methods to be used in screening assays and three approaches for confirmatory assays yielding 21 different combination of approaches. We have attempted to be as comprehensive as possible in the methods investigated, yet the sheer number of approaches currently in the literature disallowed a full evaluation. The most notable ideas that have not been considered here is the simplified decision tree in [10] and the fixed percent inhibition method [6,9]. The former was excluded as initial evaluations revealed an undistinguishable performance to the decision tree in [6] while the latters subjective choice of what percentage ought to be used was prohibitive.

In this section we will describe the different methods for classifying samples. The principle idea of each approach for confirmatory assays is to determine if the change in assay signal with and without pre-incubation of a sample with high amounts 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 without preincubation for each sample are available (the screening data) and that measurements with and without preincubation are available for confirmation. For the latter we also assume pre-incubation is successful and truly leads to inhibition. Moreover, we assume that multiple runs (analyses) per sample are undertaken and that measurements are corrected for run noise. As in [9] we will use an average of the runs per sample (e.g. mean per subject across runs) to utilize multiple runs recognizing that more involved methods may be necessary depending on the underlying experimental design (e.g. [11]). Measurements with pre-incubation of the therapeutic drug will be referred to as "inhibited measurements" and without incubation as "uninhibited measurements".

2.1. Methods for classification: Screening assays

2.1.1. Method S1: 95th percentile

The cut point is found as the 95th percentile of the uninhibited observations.

2.1.2. Method S2: Parametric method

The cut-off value is calculated as $\bar{X} + z_{0.95} * \mathrm{SD}$, where $\bar{X}$ and SD are the mean and standard deviation of the uninhibited measurements respectively and $z_{0.95}$ is the 95% percentile of the standard normal distribution (approximately 1.645).

2.1.3. Method S3: Robust parametric method

The cut point is found as $\tilde{X} + z_{0.95} * 1.483 * \mathrm{MAD}$, where $\tilde{X}$ and MAD are the median and median absolute deviation of the uninhibited measurements respectively and $z_{0.95}$ is the 95% percentile of the standard normal distribution as before.

2.1.4. Method S4: Decision tree

The following decision tree, as described in [6], is used to find the cut-point.

1. Perform a Shapiro–Wilks test [12] to assess normality of the uninhibited data. If the p-value is <0.05 the data are log-transformated.
2. Calculate the 25% and 75% percentile, $X_{0.25}$ and $X_{0.75}$, of the (transformed) data. Eliminate all data points outside the interval $[X_{0.25} - 1.5 * (X_{0.75} - X_{0.25});\ X_{0.75} + 1.5 * (X_{0.75} - X_{0.25})]$. This corresponds to eliminating data that are classed as outliers in a box–whisker plot (e.g. [13]).
3. Perform the Shapiro–Wilks test [12] to assess normality using the remaining data. If the p-value is <0.05, use the 95% percentile to calculate the intermediate cut point, otherwise the parametric method is used.
4. If data were log-transformed take the anti-logarithm of the intermediate cut point as final cut point otherwise the intermediate cut point is the final cut point.

Note, that in general it is not recommended to test every data set for normality and use the result to decide between parametric and nonparametric statistical tests (e.g. [14,15]). This procedure has, however, been proposed as a compromise between statistical rigor and practicality.

2.1.5. Method S5: Mixture model

This method, which has been proposed in [7], aims to identify if samples are negative or positive and then only uses the negative samples to find the cut point. The approach uses (regression) mixture models (e.g. [16–18]) that allow different populations (in this application positive and negative subjects) to follow different probability distributions.

The approach is to firstly identify, using the Bayesian Information Criterion (BIC) if there is more than one population in the screening data. If there is more than one population, then only samples belonging to the larger population, which is assumed to be corresponding to negative samples, will be used for cut point determination while all screening data are used otherwise. The cut point is then found as the 95th percentile of the observations. A formal description and details on the specific implementation of this method are provided in the Supplementary Material.

2.1.6. Method S6: Prediction intervals

This approach is advocated in [8] and is based on obtaining intervals for future observations based on m historical observations. In particular the cut-point is found as $\bar{X} + t_{0.95,m-1} * \mathrm{SD} * \sqrt{1 + 1/m}$, where $\bar{X}$ and SD are the mean and standard deviation of the uninhibited measurements respectively and $t_{0.95,m-1}$ is the 95% percentile of a t-distribution with $m - 1$ degrees of freedom.

2.1.7. Method S7: Experimental approach

The experimental approach, which utilizes screening and confirmatory assay data together obtains the cut point through the following steps:

1. Find a preliminary cut point for the inhibited samples based on the 95% percentile method;
2. Use the preliminary cut point to classify uninhibited values into positive and negative samples;
3. Create a new dataset containing all screening samples below the preliminary cut point and all screening samples larger than the preliminary cut-off value provided that the confirmatory value is larger than the screening value. The second set of samples is included as such observations correspond to an nonspecific signal (false positives);
4. Use the 95% percentile method with the new dataset to get the final cut-point.

2.2. Methods for classification: Confirmatory assays

2.2.1. Method C1: Parametric difference

Find the difference between uninhibited and inhibited measurement for each sample

$D =$ uninhibited measurement $-$ inhibited measurement.

The cut point is found as $c_D = \bar{D} + z_{0.999} * \sigma_D$ where $\bar{D}$ 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).