Journal of Pharmaceutical and Biomedical Analysis 55 (2011) 1148–1156

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

A formal comparison of different methods for establishing cut points to distinguish positive and negative samples in immunoassays

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

[a] Medical and Pharmaceutical Research Unit, Deparment of Mathematics and Statistics, Lancaster University, Lancaster LA14YF, United Kingdom
[b] Baxter Innovations GmbH, Wagramer Strasse 17-19, 1220 Vienna, Austria

ARTICLE INFO

Article history:
Received 9 February 2011
Received in revised form 6 April 2011
Accepted 8 April 2011
Available online 15 April 2011

Keywords:
Anti-drug antibody
Cut point
ELISA
Immunoassay
Immunogenicity

ABSTRACT

Biotechnology derived therapeutics may induce an unwanted immune response leading to the formation of anti-drug antibodies (ADA). As a result the efficacy and safety of the therapeutic protein could be impaired. Neutralizing antibodies may, for example, affect pharmacokinetics of the therapeutic protein or induce autoimmunity. Therefore a drug induced immune response is a major concern and needs to be assessed during drug development. It is therefore crucial to have assays available for the detection and characterization of ADAs. These assays are used to classify samples in positive and negative samples based on a cut point. In this manuscript we investigate the performance of established and newly developed methods to determine a cut point in immunoassays such as ELISA through simulation and analysis of real data. The different methods are found to have different advantages and disadvantages. A robust parametric approach generally resulted in very good results and can be recommended for many situations. The newly introduced method based on mixture models yields similar results to the robust parametric approach but offers some additional flexibility at the expense of higher complexity.

1. Introduction

Biotechnology derived therapeutics may induce an unwanted immune response resulting in the formation of anti-drug antibodies (ADA). As a consequence of the development of ADA efficacy and safety of the therapeutic protein could be impaired. For example, binding or neutralizing antibodies may affect pharmacokinetics or functionality of the therapeutic protein or even induce autoimmunity when the ADA cross-react with endogenous counterparts. In addition, unwanted immune responses may lead to allergic reactions. As a result, drug induced immune responses to a therapeutic protein are a major concern and need to be assessed during drug development.

Consequently there is a need to develop appropriate assays for the detection and characterization of ADA. In 2007, the European Medicines Agency (EMA) published a guideline that describes the general strategy for the development and validation of assays for immunogenicity assessment of biotechnology derived therapeutic proteins [1]. A multi-tiered approach for the testing of patient samples is recommended. In the first instance a screening assay is used for rapid identification of positive samples while subsequently an additional confirmatory assay is used to confirm the results of the screening assay. As a third step, a functional assay for assessment of the neutralizing capacity of antibodies is recommended. Screening, confirmatory and functional assays for detection and characterization of ADA need to be validated [2,3].

A critical step during assay development and validation is the definition of an appropriate cut-off that can be used to distinguish between positive and negative samples in the screening assay. This initial assay needs to be as sensitive as possible to maximize the detection of true positive samples and should be designed to avoid classifying positive samples as negative. A proportion of false positive samples is acceptable as they can be identified by the following confirmatory assay while costs and time urge to take few samples to this second stage. This approach ensures that the assays will detect as many patients who have indeed developed antibodies.

A valid statistical approach needs to be elaborated to define a reliable cut-off value used in screening and confirmatory assays [4]. For defining an appropriate cut point usually control samples obtained from healthy subjects or untreated patients are used. Such a pool of control samples is in most cases of heterogeneous composition, containing sub-populations consisting of true negative samples as well as true and false positive samples. The portion of each sub population has impact on the final cut-off value if one assumes that indeed all samples are truly negative. For example, a high content of true positives in the sample population due to specific pre-existing antibodies used for calculating the cut-off would result in a high number of false negative evaluation of samples.

* Corresponding author.
E-mail address: jaki.thomas@gmail.com (T. Jaki).

0731-7085/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2011.04.006

Consequently it is crucial to use statistical methods that deal with potential (false) positive samples appropriately when determining a cut point. Different strategies to detect and characterize ADA's have been discussed in [4,5] but no formal evaluation of the methods has yet been undertaken.

In this paper we evaluate a variety of established and less established methods for cut point determination. We will introduce the methods in Section 2 before we compare them thoroughly via simulation (Section 3). We conclude with an in-depth discussion and some future directions.

2. Methods to determine cut point

In this section we will describe various methods for determining cut points. Many of the methods are informed by the discussions in [4], although some adjustments have been made to enable automated cut point determination in the simulations to follow. Most importantly no outlier removal is incorporated prior to applying the various methods as different criteria will result in different cut points. Furthermore, the simulated data studied later do not contain outliers and hence such outlier removal will not be necessary. The conclusions made from the evaluation is nevertheless transferable to situations were outliers are present and subsequently removed. Finally note that the methods discussed here establish a fixed cut point. We will briefly highlight how one of the methods can be extended for floating cut points in the data application and the discussion.

2.1. Method 1: 95th percentile

The cut point is found as the 95th percentile of the screening data. This method does not assume a distribution of the measurements and will result in a false positive rate of 5% if indeed all samples are truely negative.

2.2. Method 2: parametric method

The cut-off value is calculated as $\bar{X} + z_{0.95} \times \mathrm{SD}$, where $\bar{X}$ and SD are the mean and standard deviation of the screening measurements respectively and $z_{0.95}$ is the 95% percentile of the standard normal distribution (approximately 1.645). This method assumes that the measurements are normally distributed. If all samples are negative and the normality assumption is satisfied, it will result in a false positive rate of ∼5%.

2.3. Method 3: robust parametric method

The cut point is found as $\tilde{X} + z_{0.95} \times 1.483 \times \mathrm{MAD}$, where $\tilde{X}$ and MAD are the median and median absolute deviation of the screening measurements respectively and $z_{0.95}$ is the 95% percentile of the standard normal distribution as before. This method resembles the parametric method but uses robust estimators of center and spread. It is designed to yield improved results if measurements are not normally distributed and similar results to the parametric method for normal data.

2.4. Method 4: decision tree

A decision tree approach is used to arrive at the cut-off value. The implementation considered here is taken from the left panel of Fig. 1 in [4] and specifically is calculated according to the following steps.

1. Perform a Shapiro–Wilks test [6] to assess normality of the screening data. If the p-value is <0.05 the data are log-transformated.

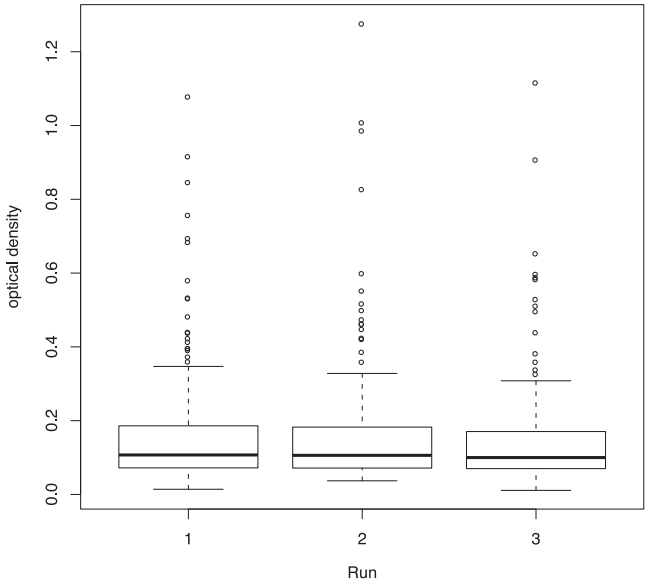

Fig. 1. Boxplot of screening values obtained in three runs of 157 healthy volunteers.

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 $[X_{0.25} - 1.5 \times (X_{0.75} - X_{0.25});\ X_{0.75} + 1.5 \times (X_{0.75} - X_{0.25})]$. This corresponds to eliminating data that are classed as outliers in a box-whisker plot (e.g. [7]).
3. Perform the Shapiro–Wilks test [6] 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.

The above algorithm aims to identify which method is most appropriate by assessing the distribution of the screening values prior to deciding which approach to take. It thereby tries to bring together the advantages of different methods by combining them which comes at the expense that the method used to find the cut point is data dependent and therefore not known a priori.

In general, however, 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. [8–10]). Decisions about when to use parametric or nonparametric tests should be made to cover an entire series of analyses. In addition, with large samples like the ones in immunoassys, minor deviations from normality may be flagged as statistically significant, even though small deviations from a normal distribution will not affect the results.

2.5. Method 5: mixture model

This method, which has not been proposed previously, aims to identify if samples are negative or positive and then only uses the negative samples to find the cut point. It employs so-called (regression) mixture models which have been shown to be useful in many scientific contexts (e.g. [11,12]). A full mathematical description of these models can, for example, be found in [13]. The idea behind such models is that different populations (in this application positive and negative subjects) are described by different probability distributions.

The use of these models here is therefore to firstly identify if there is more than one population in the screening data. If there is more than one population, then only samples belonging to the