Journal of Immunological Methods 373 (2011) 200–208

Contents lists available at SciVerse ScienceDirect

Journal of Immunological Methods

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

Computational modeling

Statistical considerations for calculation of immunogenicity screening assay cut points

David Hoffman [a,*], Marion Berger [b]

[a] Early Development Biostatistics, Sanofi-Aventis, Bridgewater, NJ, 08807-0890, United States
[b] Research & CMC Biostatistics, Sanofi-Aventis, Montpellier, France

ARTICLE INFO

Article history:
Received 26 April 2011
Received in revised form 9 August 2011
Accepted 23 August 2011
Available online 1 September 2011

Keywords:
Immunogenicity
Anti-drug antibody
Cut point
False positive rate
Prediction limit

ABSTRACT

Most therapeutic proteins induce an unwanted immune response. Antibodies elicited by these therapeutic proteins may significantly alter drug safety and efficacy, highlighting the need for the strategic assessment of immunogenicity at various stages of clinical development. Immunogenicity testing is generally conducted by a multi-tiered approach whereby patient samples are initially screened for the presence of anti-drug antibodies in a screening assay. The screening assay cut point is statistically determined by evaluation of drug-naïve samples and is typically chosen to correspond to a false positive rate of 5%. While various statistical approaches for determination of this screening cut point have been commonly adopted and described in the immunogenicity literature, the performance of these approaches has not been fully evaluated. This paper reviews various statistical approaches for cut point calculation, evaluates the impact of sampling design and variability on the performance of each statistical approach, and highlights the difference between an 'average' or 'confidence-level' cut point in order to develop more specific recommendations regarding the statistical calculation of immunogenicity screening cut points.

1. Introduction

The increasing use of biotechnology-derived proteins as therapeutic agents has highlighted the need for the strategic assessment of immunogenicity at various stages of clinical development. Most therapeutic proteins induce an unwanted immune response which may be triggered by numerous patient-, disease-, product-, or process-related factors (EMA, 2007; Jahn and Schneider, 2009). The potential consequences of such an immune reaction to a therapeutic protein range from the benign transient appearance of antibodies to life-threatening conditions (EMA, 2007; FDA, 2009; Shankar et al., 2006). Clinical consequences may include, among others, loss of efficacy, altered pharmacokinetics, administration reactions, and anaphylaxis (EMA, 2007; Shankar et al., 2006, 2008). Accordingly, the evaluation of potential immunogenicity has been a recent focus of regulatory concern (FDA, 2009; EMA, 2007, 2009; Shankar et al., 2006).

The evaluation of clinical and nonclinical immunogenicity is generally performed via detection and characterization of anti-drug antibodies. A number of different analytical formats are available for the detection of anti-drug antibodies, including direct or bridging enzyme-linked immunosorbent assays (ELISA), radioimmunoprecipitation assays (RIPA), surface plasmon resonance (SPR), and electrochemiluminescence assays (ECL) (FDA, 2009). While each format has relative advantages and disadvantages, consideration should be given to product characteristics, potential co-medications, disease-specific issues, and epitope exposure when selecting a format (Mire-Sluis et al., 2004). Regardless of the chosen format, the assay must be validated for its intended purpose. Such validation generally includes assessment of linearity, accuracy, precision, specificity, selectivity, stability, detection and/or quantification

* Corresponding author. Tel.: +1 908 981 6287.
E-mail addresses: david.hoffman@sanofi-aventis.com (D. Hoffman), marion.berger@sanofi-aventis.com (M. Berger).

0022-1759/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jim.2011.08.019

limits, robustness, and system suitability (FDA, 1999; Mire-Sluis et al., 2004; ICH, 1996).

Immunogenicity testing is generally conducted by a multi-tiered approach whereby patient samples are initially screened for the presence of anti-drug antibodies in a screening assay (Koren et al., 2008). Samples testing positive for the presence of anti-drug antibodies in the screening assay are subsequently analyzed in a confirmatory assay which characterizes the specificity of the binding response to the drug. Samples confirmed for the presence of anti-drug antibodies are then typically analyzed in a neutralizing antibody (NAB) assay to assess the neutralizing capacity of the anti-drug antibodies. The screening assay is intended to provide a rapid and sensitive initial assessment of the samples, while the confirmatory and neutralizing antibody assays are generally more labor-intensive and time-consuming.

In this multi-tiered approach to immunogenicity assessment, a key consideration is the determination of a screening assay cut point. The screening assay cut point is the level of response of the screening assay at or above which a sample is defined to be positive for the presence of anti-drug antibodies and below which it is defined to be negative. This cut point should be statistically determined by evaluating samples deemed to be representative of the drug-naïve target subject/patient population (i.e. negative control samples) (FDA, 2009). Typically, the cut point is chosen to correspond to a false positive rate of 5% (FDA, 2009). This is intended to control the number of unnecessary confirmatory assays while providing some assurance that the false negative rate will be small. However, it should be noted that the false negative rate cannot be determined from the false positive rate and can only be estimated from evaluation of samples deemed to be representative of the target population of patients/subjects with drug-induced immune response (i.e. positive control samples).

Several statistical approaches for determination of the screening cut point are commonly used and have been described in the immunogenicity literature (Mire-Sluis et al., 2004; Gupta et al., 2007; Shankar et al., 2008; FDA, 2009). These may include parametric, robust parametric, or non-parametric approaches. There has been some recent investigation of the performance characteristics of various statistical approaches for cut point determination (Schlain et al., 2010; Jaki et al., 2011). Further, some general recommendations regarding the number of samples and independent assay runs used to determine the assay cut point have been proposed (Gupta et al., 2007; Shankar et al., 2008; FDA, 2009). However, there has been little consideration of the required control of the nominal false positive rate necessary to meet regulatory expectations and achieve suitable analytical performance. Moreover, the impact of the sampling design (i.e. number of subject/patients and assay runs) and relevant sources of variability on cut point performance characteristics has not been thoroughly explored. The purpose of this paper is to review various statistical approaches for cut point calculation, evaluate the impact of sampling design and variability on the performance of each statistical approach, and highlight the difference between an 'average' or 'confidence-level' cut point in order to develop more specific recommendations regarding the statistical calculation of immunogenicity screening cut points.

2. Methods

2.1. Sampling design

The importance of proper experimental design when determining a screening cut point has been recognized in the immunogenicity literature (Shankar et al., 2008). Multiple factors may introduce variability into the observed assay responses. Such factors may include the assay run (or batch), analyst, plate, and plate location, among others. Care should be taken to utilize an experimental design which allows for the direct evaluation of the effect of each factor, eliminating potential confounding between the effects of two or more individual factors. See Shankar et al. (2008) for a brief discussion of balanced experimental designs to eliminate such confounding.

The chosen experimental design will determine which sources of variability are directly estimable from the observed assay responses, and the calculated cut point should incorporate each relevant source. At minimum, three sources of variability should be directly estimable from the observed assay responses and incorporated into the calculated cut point:

- Biological inter-subject (or inter-patient) variance
- Analytical inter-run (or inter-batch) variance
- Analytical intra-run (or intra-batch) variance

Depending on the assay format and experimental design, additional sources of variability may be relevant and identifiable. For simplicity, we will consider only the above three sources of variability for the remainder of this paper. Assuming all samples are assayed in each analytical run, a statistical model to describe the assay responses can then be given by:

$$y_{ij} = \mu + s_i + r_j + \varepsilon_{ij} \quad (1)$$

where y_{ij} is the observed assay response for the ith ($i = 1,2,\ldots,I$) subject in the jth ($j = 1,2,\ldots,J$) assay run, μ is the true (unknown) mean response for the assay, s_i is the random effect for the ith subject, r_j is the random effect for the jth assay run, and ε_{ij} is the random effect for the ith subject in the jth assay run. Under the assumption that the observed assay responses follow a normal distribution (or that a transformation to achieve normality exists), we can further specify that the random effects s_i, r_j and ε_{ij} are normally and independently distributed with means zero and variances σ_S^2, σ_R^2 and σ_E^2. These variances, σ_S^2, σ_R^2 and σ_E^2, correspond to the biological inter-subject, analytical inter-run, and analytical intra-run variability, respectively. The total variability of an observed assay response is then given by $\sigma_y^2 = \sigma_S^2 + \sigma_R^2 + \sigma_E^2$.

The statistical model given in Eq. (1) is commonly referred to as a balanced two-way random effects model (without interaction). Interaction terms or additional random effects may be incorporated into the analysis model as appropriate based on the experimental design and assay format. Denote the overall mean of the observed assay responses by $\bar{y} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}/IJ$, the mean for the ith subject by $y_i = \sum_{j=1}^{J} y_{ij}/J$, and the mean for the jth assay run by $y_j = \sum_{i=1}^{I} y_{ij}/I$. Table 1 gives the analysis of variance table for the balanced two-way random effects model, where EMS denotes the expected mean square.

The mean squares MS_S, MS_R, and MS_E can be used to obtain estimates of the inter-subject, inter-run, intra-run, and