The AAPS Journal (2019) 21: 55
DOI: 10.1208/s12248-019-0328-8

Meeting Report

Report on the AAPS Immunogenicity Guidance Forum

Heather Myler,[1,9] Boris Gorovits,[2] Kelli Phillips,[1] Viswanath Devanarayan,[3] Adrienne Clements-Egan,[4] George R. Gunn,[5] Susan Kirshner,[6] Binodh DeSilva,[7] and Vinod P. Shah[8]

Received 4 March 2019; accepted 3 April 2019; published online 16 April 2019

Abstract. In September 2018, the American Association of Pharmaceutical Scientists (AAPS) conducted an Annual Guidance Forum on the considerations related to immunogenicity testing for therapeutic protein products. In addition to a broad representation by the pharmaceutical industry, the event included strong representation by leading scientists from the US Food and Drug Administration (FDA). The agency and industry perspectives and updates to the guidance were presented. Specific topics that were discussed included the strategies of anti-drug antibody (ADA) assay cut-point assessments, the selection of ADA-positive controls (PCs), and the evaluation of PC performance. Assessment strategies and relevance of ADA assay attributes were also discussed, including assay drug tolerance and ADA assay sensitivity. The following is a summary of the discussion.

KEY WORDS: Immunogenicity; Anti-drug antibodies (ADA); Neutralizing antibodies (NAb); FDA guidance.

INTRODUCTION

Our understanding of ADA assays has greatly evolved over the last two decades, starting with the Mire-Sluis paper published in 2004 describing the design and optimization of immunoassays in the detection of host antibodies against biotherapeutic products, followed by the Shankar paper in 2008 addressing associated validations, the Korean paper in 2008 addressing risk-based strategies and many papers thereafter, culminating in guidance documents issued by the FDA, European Medicines Agency (EMA), Ministry of Health, Labour and Welfare (MHLW), and various other health authorities (1–5). Discussions provided on the draft guidance in this report are consistent with the final guidance issued in January 2019.

In general, the FDA guidance is a good reflection of the most common immunogenicity assay development and validation practices applied in the biopharmaceutical industry. However, there are a number of recommended assay attributes that are still debated, such as the rationale for producing extremely sensitive ADA assays, the process of determining low-screen (and confirmatory) cut-point values, the relevance of pre-existing Abs, and the clinical relevance of ADAs detected in these highly sensitive assays.

Low assay signal variability across samples from individuals frequently results in impractical screen and confirmatory cut-points. ADA assay platforms and methodologies have matured, frequently resulting in low analytical and biological variability. Despite the use of several operators and hundreds of data points from a set of validation runs, ADA assays may have very low variability that results in extremely low cut-points that do not stand the test of time across studies. This sometimes, though not always, results in a false positive ADA-screening incidence well above 5%. We also continue to ask ourselves about the accuracy of a confirmed ADA-positive result when we have a very low specificity cut-point, particularly when it is accompanied by a small raw signal shift such as 10 ECL units.

Additionally, assays are frequently much more sensitive than the 100-ng/mL ADA level which is thought to be potentially clinically meaningful and recommended in the guidance. While the guidance has the good intention to confirm that the assay sensitivity is maintained by ensuring the low positive control (LPC) hovers near the cut-point and has a 1% failure rate, many workshop attendees had concerns that this stringent LPC 1% failure requirement would not add value to ensure detection of clinically meaningful ADA. This

[1] Immunochemistry Department, PPD Laboratories, 2244 Dabney Road, Richmond, Virginia 23230-3323, USA.
[2] Biomedicine Design, Pfizer, Andover, Massachusetts 01810, USA.
[3] Statistics and Data Sciences, Charles River Laboratories, Horsham, Pennsylvania 19044, USA.
[4] Biologics Development Sciences, Janssen, Spring House, Pennsylvania 19477, USA.
[5] Immunogenicity and Clinical Immunology, GSK, Collegeville, Pennsylvania 19426, USA.
[6] Product Quality and Immunogenicity, FDA, Silver Spring, Maryland 20903, USA.
[7] Analytical Strategy & Operations, Bristol-Myers Squibb, Princeton, New Jersey 08540, USA.
[8] Pharmaceutical Consultant, North Potomac, Maryland 20878, USA.
[9] To whom correspondence should be addressed. (e–mail: Heather.Myler@ppdi.com)

is because the actual lowest concentration of ADA detected in the assay (which is often < 100 ng/mL) may be 10 or more fold lower than the steady state drug trough level (e.g., 1–100 μg/mL). In other words, the assays are already extremely sensitive and are likely to detect both clinically meaningful and irrelevant ADAs without adding a stringent LPC 1% failure requirement. Attendees at the workshop questioned whether there are other suitable ways to demonstrate appropriate assay performance over time that would not result in unnecessary QC failures during study phase bioanalysis.

INDUSTRY AND HEALTH AUTHORITY PERSPECTIVES AND BEST PRACTICES

Health authorities (HAs) consider immunogenicity assays as threshold assays, designed to detect the ADA concentrations known to trigger life-threatening implications for patients prescribed biotherapeutics. For pre-clinical analysis and first-in-human studies, ADA methods may be in a preliminary state. However, ADA methods are expected to improve as drug development progresses to late phase. For Phase III analysis, ADA methods are expected to be fully validated with well-controlled assay acceptance criteria and sufficient characterization data to support results from patients who are undergoing treatment for the targeted disease indication (5).

Screening Cut-Point (SCP) and Confirmatory Cut-Point (CCP) Determination

Health Authority Perspective. The strategy for calculation of a statistically derived cut-point remains a primary focus for regulatory review of ADA methods. While HAs may consider innovative approaches for pre-study cut-point assessment to be acceptable, the original process for calculating ADA assay cut-points that was presented in 2008 remains standard (2). Only minor modifications to this approach have been implemented as technology advances and the industry understanding evolves.

The initial recommendation for the calculation of the screening cut-point, designed to identify potentially positive ADA samples with a false positive rate (FPR) of 5%, uses the mean of individual matrix responses and the standard deviation (SD) of the responses, multiplied by a factor of 1.645 that corresponds to the 95% percentile of a normal distribution. For each subsequent tier of analysis, the targeted FPR decreases. The recommended FPR for confirmatory assays is 1%. Following a positive response in the confirmatory assay, the characterization of the ADA response (i.e., neutralizing antibody assays) also may target a 1% FPR. ADA response magnitude, evaluated by diluting positive samples in a series until the response falls below the assay threshold (titer cut-point), may be calculated targeting only a 0.1% FPR. As the above formula for calculating the cut-points are only point estimates of the percentiles, they provide only 50% confidence for achieving the targeted false positive rate. Statistical formulae for calculating the cut-points that provide higher confidence for achieving the targeted false positive rates were presented in 2015 (6) and stated in the 2016 draft FDA guidance. In addition, the use of median and 1.4826 times the median absolute deviation (MAD) instead of mean and SD of individual matrix responses (*vs.* the mean) to account for bias for potential borderline outlier samples has been recommended as an alternative strategic approach for achieving reliable cut-points (6). The median/MAD approach is offered as an alternative and is not intended to discourage the use of other robust statistical approaches that employ mean/SD.

Industry Perspective. The FDA guidance states that cut-point should be determined statistically with an appropriate number of treatment-naïve samples, generally around 50, from the subject population and that each sample should be tested by at least two analysts on at least three different days for a total of at least six individual measurements. In practice, however, pre-study validation cut-point data are supplemented using in-study pre-dose sample data and the targeted 5% FPR. Given this supplemental assessment, many participants feel that six individual measurements may not be needed and that four may be sufficient during pre-study validation. Further, if the target 5% FPR is not met, then, in-study cut-points are assessed using pre-dose samples run either once or at the most twice. While the industry agrees with sound statistical approaches to cut-point assessments, there are times when it may not be possible to have a perfect balanced design for cut-point assessments, particularly when the assay is supporting a rare population or when an in-study cut-point is needed. For most assessments, > 50 individuals are recommended. However, in instances when an in-study cut-point will be evaluated prior to sample analysis or there are fewer than 50 patients in a given clinical trial, fewer individuals may be used ($n = > 20$).

Positive Control Source

Health Authority Perspective. ADA assays are characterized and later monitored using a positive control sample that is prepared by combining ADA with the negative control (naïve matrix pool developed/analyzed during cut-point assessment). Assay sensitivity that is characterized using positive controls that have not been purified may inaccurately indicate the assay has poor sensitivity. Additionally, matrix components from an immunized animal may impact the positive control response. Therefore, HAs have recommended that positive controls are purified from source matrix and have a known concentration. Positive controls that are monoclonal (developed against a single epitope of the drug molecule) and/or polyclonal (developed against the whole drug molecule) may be used during method development and validation. Polyclonal ADAs best represent an endogenous ADA response and are recommended to be used for assay characterization experiments (selectivity, dilutional linearity, drug tolerance, and hemolysis/lipemia recovery). Monoclonal ADAs may have a higher affinity and may be used to prepare robust positive controls used to monitor assay performance.

Industry Perspective. Industry practice is largely in line with HA recommendations. Further, it is recommended to

provide details of the positive control source and specificity in validation reports and filings for added clarity.

Sensitivity

Health authority perspective. It is well understood by HAs that the characterization of assay sensitivity is dependent upon the affinity of the positive control and may not be representative of the patient ADA response. Additionally, historic characterization of assay sensitivity has not routinely included a discussion of assay drug tolerance and impact of the disease state serum or long-term method use. Therefore, assay sensitivity is best understood by considering a range of values, rather than fixed values that are based on a single-source control. FDA guidance released in January 2019 recommends improving the target for assay sensitivity from 250 to 100 ng/mL in a neat matrix. The guidance also recommends a LPC set to fail 1% of the time based upon the screen and confirm tier cut-points noting that this is more important for high-risk programs and less important for low-risk programs. The guidance also indicates that assay sensitivity must be defined for both the assay screening and confirmatory tiers. The two assays are expected to demonstrate comparable sensitivity, thought to be impacted most significantly by the increased confidence interval (99%) used to calculate the confirmatory cut-point.

Industry Perspective. The guidance states that the LPC should be set at levels where 1% of the LPC should screen/confirm negative. In many instances, LPC concentrations are far below the 100 ng/mL recommended sensitivity so the added value of this recommendation to ensure an adequately sensitive ADA assay over time was questioned. In addition, determining an LPC concentration that meets this 1% failure criterion during pre-study validation is challenging and could result in the need to re-determine the LPC as the assay is used over time. One option could be to set the LPC slightly higher in initial clinical studies and refine the LPC level over time to adequately cover ADA expected in Phase III studies. While it is possible to adjust the LPC over time to meet this goal, the continual assay PC adjustments and unnecessary plate failures due to this criterion lead to inefficiency and sample waste.

Drug Tolerance

Health Authority Perspective. HAs acknowledged that well-characterized ADA method drug tolerance may be the most important parameter needed to fully understand ADA response *in vivo*. Samples collected prior to the therapeutic washout period may have circulating drug that has the potential to mask positive results or have a significant impact on the magnitude of detected ADA. HAs recommend testing assay drug tolerance using a range of positive control concentrations, as well as a range of drug. Concentrations of drug selected for evaluation should target concentrations that are at lower-than-expected, at expected, and at higher-than-expected levels for sample trough. PC should be tested at the LPC concentrations, 100, 250, and 500 ng/mL, and potentially higher concentrations. Drug tolerance should be evaluated in both the screening and confirmatory tiers. Verification of drug tolerance is required by assessing drug tolerance in at least two runs using the final method.

Industry Perspective. Drug tolerance samples should be prepared and stored in a method that will most closely mimic sample conditions. For example, drug and ADA should incubate together at room temperature (or 37°C) for at least an hour to allow ADA/drug interactions to occur. In addition, prepared samples should have at least one freeze thaw cycle prior to analysis. Industry suggests an increased value in confirming that drug tolerance results are reproducible. An increased number of drug tolerance sample preparations/replicates ($n = > 3$) may be evaluated during the assay validation. As well, industry suggests that drug tolerance data is presented to concisely show the ADA/PC concentrations and results for each condition evaluated.

Selectivity

Health Authority Perspective. HA have reiterated the need for thorough characterization of assay selectivity. This includes testing the recovery of a low concentration of ADA (often equal to the selected concentration of the LPC) in normal and disease state individual matrix, as well as in hemolytic and lipemic serum. Selectivity assessments also may extend to recovery of ADA in the presence of structurally similar concomitantly dosed therapeutics and other matrix factors including therapeutic target and rheumatoid factor (RF) that may impact ADA recovery. If the therapeutic shares similarity with proteins that are endogenous in the patient population, cross-reactivity of the native human protein also should be evaluated in the method.

Industry Perspective. Industry practice is largely in line with HA recommendations although there still seems to be a lack of consensus concerning interference testing using hemolytic and lipemic matrices. Explorations of the impact of hemolytic and lipemic matrices during assay development will likely be useful in determining the susceptibility of methods to such factors. Additionally, it was agreed upon by industry and health authority representatives that selectivity may be tested at levels higher than the LPC (in the 1% failure range based upon cut-point) when selectivity cannot be demonstrated at the LPC. Sensitivity does not need to be re-established based upon acceptable selectivity at the lowest control concentration but should be reported accordingly in the selectivity section of the validation report.

In practice, selectivity is often performed with and reported for PC. Samples from at least 10 individuals, per disease state, are tested as blank matrix and matrix spiked at LPC and HPC levels. It is standard practice to include spike controls (LPC and HPC) in the pooled matrix so selectivity samples can be evaluated in the context of selectivity, minimizing the question of random QC bias or spike error. Selectivity is performed in both the screen and confirmatory

tiers and acceptance is based on the screen and confirm cut-points. We do not recommend calculating the percent difference from control response as a means of assessment. Instead, it is recommended to use the established acceptance criteria for PCs. Blank matrix should be < CP and have percent inhibition < CCP to verify as negative. LPC should be > CP and have percent inhibition > CCP to verify as positive. In cases where the LPC has a lower bound other than the CP, that lower bound should be used in place of the CP. HPC should be > lower bound based upon a 99% lower limit and have percent inhibition > CCP. Selectivity is generally considered acceptable if 80% of the individuals tested meet criteria. However, cases do exist where a high prevalence of reactivity is observed, which could be attributed to a high prevalence of pre-existing antibodies. Health authorities did note that it is acceptable to select selectivity samples from individuals understood to be negative per cut-point experiments. Pre-existing reactivity should be reported to understand the prevalence of potential pre-existing antibodies.

In addition to selectivity testing in individuals as described above, specificity can be tested with and without potential interferant(s), including those mentioned previously (concomitant therapeutics, target, RF, etc.). This testing in normally performed in the pooled matrix.

Stability

Health authority perspective. Health authority expectations and industry practice seem to be well aligned for stability analysis. Both agree that long-term frozen storage stability evaluation of the surrogate PC is unnecessary in accordance with Shankar *et al*., (2) which cites the stability of antibodies for 2 years or longer when stored at temperatures ≤ − 20°C. Freeze/thaw stability including at least five freeze thaw cycles and process temperature stability, 4–24 h at ambient room temperature and/or 4°C, depending on assay conditions, is considered necessary. This stability testing is performed using the NC, LPC, and HPC tested in both the screen and confirmatory assay tiers. As with selectivity, it is not necessary to calculate the percent deviation from control responses to assess sample stability. It is suggested to instead use the established acceptance criteria for each control as described previously.

Parameters that May Be Evaluated During Assay Development

Health Authority Perspective. Assay development data may be used to support assay validation during ADA method review by HAs. Parameters that often are evaluated early in method development that may not need to be repeated during method validation include determination of the minimum required dilution (MRD) and assay positional effects. Development data also may be used to support characterization of the assay robustness and critical reagent stability. Some of the more common robustness parameters frequently examined by industry practitioners in method development and/or validation include inter-analyst variability (MD and MV), plate lots, reagent lots, incubation times, and equipment.

Industry Perspective. Industry practice is largely in line with HA recommendations. In addition to those parameters discussed above, drug tolerance data from method development is often used to supplement drug tolerance data from validation provided the final method was used.

NAb Assays

Health Authority Perspective. The most significant ADA responses to biologics include those ADA responses that act to neutralize the function of the drug. This typically occurs when the ADA block the interaction of the drug and target. Characterization of an ADA impact on drug-target interaction is assessed using neutralizing antibody assays (NAb Assays). NAb assays may be competitive ligand binding-based or cell-based depending on the immunogenicity risk and therapeutic mode of action. HAs have previously noted that cell-based assays may provide a milieu that is most similar to the *in vivo* environment and this is especially important if the therapeutic target is a cell-surface receptor. The use of cell lines in the detection of neutralizing antibodies ensures that the target protein maintains accurate tertiary structure during analysis. The receptors exposed on the cell surface are available to the target protein and epitopes are available to ADA in the endogenous state. While all method development considerations were not discussed in detail, it was reiterated that cell-based assay method development documentation should include supporting data for the selected concentration of stimulatory drug to ensure that the assay is sensitive to the response inhibition that is triggered by NAb-positive samples. Further, robustness should be thoroughly evaluated with respect to cell passage number and incubation times.

Like standard immunogenicity assays, NAb assays are considered to be qualitative. The NAb assay cut-point is recommended to be based on at least 30 individual matrix responses. A confirmatory tier is not often included for NAb assays since the ADA response has already been confirmed in the standard ADA assessment. Further, NAb-positive samples are rarely tittered. HA recommends targeting assay sensitivity around 100 ng/mL. However, because of the intrinsic nature of cell-based assays, it is understood that this may not be achievable. In these cases, 250–500 ng/mL sensitivity may be sufficient, understanding that not all cell-based assays will be able to meet these criteria. In cases where 250–500 ng/mL sensitivity is not met, the HA should be consulted. This variability can also impact assay response variance between replicates. In these cases, HA recommends increasing the number of replicates tested.

Industry Perspective. It was agreed that the expectations for ADA assays generally translate to NAb assays realizing that NAb assays and particularly cell-based NAb

assays are typically more variable, less sensitive, and less drug tolerant by nature.

Post-Validation Method Evaluation

Health Authority Perspective. Method parameters should continue to be evaluated following method validation. Most significantly, the function of the validation cut-points should be confirmed using data from the study population. The false positive rate of the assay should target 5% and fall between 2 and 11% (7). LPC concentrations, determined during the validation, should have responses that fall at the low-response range of the assay to monitor performance of each run near the cut-point. The LPC is expected to fail at a rate of 1% during sample analysis. Further, HPC should cover the study data at the high end of the assay to confirm the absence of a response hook effect and to monitor precision.

Industry Perspective. Industry practice is largely in line with HA recommendations; however, debate remains concerning the 1% failure of the LPC. Industry agrees that PC should be set at levels that provide confidence in the clinically relevant ADA-positive study sample responses.

Assay Comparability

Health Authority Perspective. Assay comparability was not specifically addressed by HA at the workshop but is covered in the guidance on ADA validations.

Industry Perspective. It is general practice to perform some type of assay comparability when an assay is transferred to another lab. Minimally, sensitivity and drug tolerance are expected to be comparable through a partial validation. Another level of comparability can be assessed using individual study samples or PC with varying levels of ADA of the appropriate disease state. Although it is not necessary if the original cut-point has yielded the same results across labs, CP can be assessed at each lab.

Sample Collection Times

Health Authority Perspective. Sample collection times should be chosen to maximize the utility of data garnered from study samples. Recommended collection times are baseline, days 8, 15, 28, 45, 60, and 90, and then every 3 months while on drug (8). ADA samples should also be collected during the trough and therapeutic washout period following study completion and 30 days after last dose. If the drug is an immune suppressant, an ADA sample should be collected after a time period equal to five drug half-lives. High-risk programs may also benefit from patient consent for long-term ADA follow-up.

Industry Perspective. Industry practice is largely in line with HA recommendations. Any deviation from these recommendations should be scientifically justified and discussed with HA. It is recommended that these samples are collected should testing be merited.

DISCUSSION (BREAK-OUT SESSIONS)

Cut-Points (Screening, Confirmatory, and Titer)

The evaluation of cut-point data was discussed, and it was agreed that data should be both analyzed as pooled and on an individual run basis. The reason to conduct analysis of individual runs is to determine whether there is an alignment between analysts/days and evaluate whether all runs can be used in the determination of the cut-point. In addition, it was argued that samples that consistently appeared positive in assays run by different technicians and on different days were more likely to be due to biological variability than technical variants. The benefit of pooling data across runs is to obtain a large data set that allows for the inclusion of various analytical sources such as the inter-run, inter-analyst, inter-plate, and intra-plate variability, which will be seen during study sample analysis. Assessing the assay/analytical and biological variability, along with the identification of analytical and biological outliers, are critical for the calculation of the cut-points from pooled data.

Outlier Assessments

Outlier assessment approaches that lead to excessive outlier removal were discussed. Low CPs that are generated due to excessive outlier removal could increase the risk for reporting false positive samples that have no clinical relevance. In current practice, the use of robust statistical measures such as the median and 1.4826 times the median absolute deviation (MAD) instead of mean and standard deviation (SD) should circumvent the need for arbitrarily reducing or increasing the stringency of outlier criteria (7,9). Outlier removal should be done in a single step. Multiple iterative outlier removals should be avoided and are considered unnecessary when robust measures such as median and MAD are used to reduce the impact of borderline outliers. Assay evolution and decreasing both analytical and biological variability are the primary factors driving low cut-points and further outlier removal is not believed to mitigate this.

Relevance of Low ADA-Positive Samples

Concerns around the lack of clinical relevance of low ADA-positive samples due to highly sensitive methods and compounded by low cut-points were discussed. There was some debate about whether cut-points should be set around a clinical relevance threshold, that is a point at which ADA and Nab-positive samples are found to have impact on clinical endpoints such as exposure, safety, and efficacy, rather than on the 5% false positive rate. It was agreed that designating an ADA-positive/negative status based on the clinical

relevance needs more discussion and a suitable practical approach. It was also understood that the way the variability is calculated during pre-study validation may not fully represent all the study conditions.

Pre-Study Cut-Point Validation and In-Study Cut-Points

Pre-study validation is often done in a much more controlled and rigorous environment, and the total variability may not fully reflect the long-term practical variability that is typically encountered during the in-study phase, especially in later phase and large clinical trials. Most participants assess the FPR of pre-treatment samples from the clinical studies and determine whether it is within the 2–11% range (7) or some other pre-specified range and apply in-study cut-points only when the FPR is outside the pre-specified range. When it is determined that an in-study cut-point is needed, participants typically test all the baseline study samples only once when there are at least 50 pre-dose study samples or a minimum number of pre-dose study samples determined relevant by a statistician. Samples may be tested more than once when there are a low number of pre-dose study samples. It is important to ensure the samples are spread out across multiple assay plates and runs with more than one analyst involved. Some participants formally compare the distribution (means and variances) of the pre-study validation samples *versus* the in-study baseline/pre-treatment samples via statistical tests and visualize the data via boxplots. This is typically done to get a better understanding of the differences in the data, and sometimes to support whether in-study cut-points should be applied.

Titer Cut-Point Assessments

Titer cut-points (TCP) were discussed extensively where most participants believed the screening cut-point is often too low for reliably calculating titers. The 99.9% upper limit from the same data used to calculate SCP is typically used for calculating the TCP. Some participants use an alternative approach based on the mean + 3.09 × SD of the assay diluent. The use of titer precision as defined by the minimum significant ratio (MSR) as a guide for determining the appropriate threshold for TCP (95% *vs*. 99% *vs*. 99.9% etc.) was also discussed where the lowest threshold that provides adequate MSR (e.g., < 2) may be used for defining the TCP (7). About half of the participants in the group use a buffer to titrate the samples, while the other half use neat serum for titration. Most participants use endpoint titers instead of interpolation for reporting the titers of study samples and most participant factor in the MRD for titer calculation where the MRD is defined in terms of total sample dilution. Additionally, most participants use a threefold dilution series for calculating titers while some use a twofold dilution series. All of these approaches are deemed acceptable.

Titers Versus Screening Assay Signals

The use of signal to negative control (S/N) ratios from the screening assay as an alternative to titers was also discussed. Concerns about the top plateau and hook effect were expressed as potential risks associated with the use of S/N, acknowledging that this risk may be more relevant for the ELISAs than the ECL format. Most participants believed that a strong correlation between the S/N results and titers should be demonstrated for at least one of the clinical studies in order to support the use of S/N instead of titers for the other clinical studies involving the same therapeutic. It was widely acknowledged that this topic requires further research/investigation.

Screening Cut-Point Calculations

The use of a lower confidence limit (LCL) approach (6) for setting the SCP and CCP was discussed extensively. Participants who have tried both the LCL formula and the usual point estimate formula (2) thought that the cut-point values were practically similar for most assays. Most participants do not use the LCL formula for their clinical studies but continue to use the point estimate formula such as the one from Shankar *et al.* (2).

Several participants considered the 5% FPR to already be too conservative and requiring 90% confidence for achieving this 5% FPR *via* the use of the LCL formula to be potentially excessive. In relation to this, several participants believed that the lack of clinical relevance of the low ADA-positive samples due to the cut-points often being quite low is already a major challenge, so lowering the cut-point further *via* the use of LCL formula is often not found to be suitable. It was also discussed that the more important determinant regarding the appropriateness of the cut-points from pre-study validation should be whether the false positive rate of the pre-treatment clinical study samples is within a reasonable range around 5%, as discussed above. Susan Kirshner did mention in her presentation that there was flexibility in determining the most suitable statistics for any given program.

Confirmatory Cut-Points

The change of CCP from a 99.9 to a 99% threshold (1% false positive rate) was discussed from a practical and scientific standpoint (5). Most participants expressed that they have had to go back to their past studies and reassess the CCP according to the 99% limit, which has resulted in considerable additional work, with often no material consequence to the ADA results or clinical data interpretation. Further, the more frequently observed low CCP (e.g., < 20%) are thought to increase the risk for over-reporting false positives.

Clinical Trials, Multiple Labs, and Rare Disease

Large international trials with multiple testing facilities *versus* smaller trials for rare populations were discussed. It was agreed that monitoring for false ADA positives or negatives is important when moving to new disease or geographical populations. Re-validating assay cut-points for new populations, especially if the false positive rate is outside the pre-specified acceptance range, is needed to ensure that applicable cut-points are being used across studies, indications, and labs. Ongoing monitoring of in-study performance is important when determining whether there is a need for an

in-study CP. Further, due to the unique clinical protocol designs and potential safety implications, it was agreed that assays need to be validated early on in support of rare populations.

Monitoring Assay Performance

PC-prepared material is used to monitor assay performance, including LPC and HPC. Additional controls may be needed based on specifics of assay performance. Reporting assay performance in study reports is required. This may include raw assay signals or S/N as applicable and inter-run precision of the controls. It was agreed that performance of the LPC is typically more critical than that of the HPC when determining suitability of the assay. It may be practical to assess more than one LPC concentration during assay validation, particularly if the LPC statistically calculated to fail, that is, screen and confirm negative, 1% of the time does not align with sensitivity observed in drug tolerance and selectivity experiments. There could be a reason to run a higher LPC, with a 99% lower limit not based upon the cut-point, in these cases especially in the case of an ultra-sensitive assay where sensitivity is in the low or single ng/mL range, well under the 100-ng/mL threshold that is generally considered to potentially correlate with clinical effect. Further, if the product is not considered to be high risk, additional flexibility in defining LPC may be considered. LPC is still expected to serve as the assay suitability control and to demonstrate that the assay continues to maintain appropriate sensitivity.

CONCLUSIONS

As our understanding of assays designed to assess immunogenicity of biotherapeutics advances, the agency and industry continue a discussion on appropriate assay quality criteria. A number of key papers and guideline documents have been issued in the past decade to address questions related to ADA and NAb assay design and performance verification. Current guidelines, including EMA and FDA guidance, have a clear description of critical elements required during assay development and validation. The AAPS-FDA workshop conducted in 2018 continues the trend of an open and transparent dialogue between the agency and industry colleagues allowing further understanding of the important topic of immunogenicity monitoring for biotherapeutics. While the industry and the agencies are aligned on most topics, the following points of debate were discussed during this workshop and will require even further discussions for alignment:

- The clinical relevance of ADAs in assays with sensitivity < 100 ng/mL and associated LPC criteria.
- The clinical relevance and scientific value of ultra-low screen and confirmatory cut-point values, for e.g., < 20% difference from the NC.
- Most participants continue to use the point estimate formula for the screening cut-point with outlier analysis (2) as it commonly yields the targeted false positive rate during sample analysis (rather than the more stringent formula included in the guidance).
- Flexibility between the number of donors evaluated and the number of independent repeats/runs of each sample for determining the cut-points.
- Outlier exclusion was not specifically addressed by health authorities but there still appears to be a strong industry concern and debate.
- The value of selectivity testing in hemolytic and lipemic samples.
- What constitutes an acceptable level of agreement across assays during comparability testing.
- If buffer *versus* serum comparisons are done during method development as part of MRD determination, they do not need to be repeated in validation.

REFERENCES

1. Mire-Sluis AR, Barrett YC, Devanarayan V, Koren E, Liu H, Maia M, et al. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. J Immunol Methods. 2004;289(1–2):1–16.
2. Shankar G, Devanarayan V, Amaravadi L, Barrett YC, Bowsher R, Finco-Kent D, et al. Recommendations for the validation of immunoassays used for detection of host antibodies against biotechnology products. J Pharm Biomed Anal. 2008;48(5):1267–81. https://doi.org/10.1016/j.jpba.2008.09.020.
3. Koren E, Smith HW, Shores E, Shankar G, Finco-Kent D, Rup B, et al. Recommendations on risk-based strategies for detection and characterization of antibodies against biotechnology products. J Immunol Methods. 2008;333(1–2):1–9. https://doi.org/10.1016/j.jim.2008.01.001.
4. Committee for Medicinal Products for Human Use C. Guideline on immunogenicity assessment of therapeutic proteins 2017; Available from: https://www.ema.europa.eu/documents/scientific-guideline/guideline-immunogenicity-assessment-therapeutic-proteins-revision-1_en.pdf.
5. FDA. Immunogenicity testing of therapeutic protein products — developing and validating assays for anti-drug antibody detection. Guidance for industry. U.S. Department of Health and Human Services Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER) 2019 [cited 2019]; January 2019:[Available from: https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm629728.pdf.
6. Shen M, Dong X, Tsong Y. Statistical evaluation of several methods for cut-point determination of immunogenicity screening assay. J Biopharm Stat. 2015;25(2):269–79. https://doi.org/10.1080/10543406.2014.979196.
7. Devanarayan V, Smith WC, Brunelle RL, Seger ME, Krug K, Bowsher RR. Recommendations for systematic statistical computation of immunogenicity cut points. AAPS J. 2017;19(5):1487–98. https://doi.org/10.1208/s12248-017-0107-3.
8. Shankar G, Arkin S, Cocea L, Devanarayan V, Kirshner S, Kromminga A, et al. Assessment and reporting of the clinical immunogenicity of therapeutic proteins and peptides-harmonized terminology and tactical recommendations. AAPS J. 2014;16(4):658–73. https://doi.org/10.1208/s12248-014-9599-2.
9. U.S. Pharmacopeia and the National Formulary General Chapter <1106> Immunogenicity assays—design and validation of immunoassays to detect anti-drug antibodies. 2014.

Publisher's Note