Interpretation of Immunogenicity Results and Evaluation of Clinical Associations

V. Devanarayan, Ph.D.
AbbVie, Inc.

European Bioanalysis Forum (EBF) 9th Open Meeting: Reaching Utopia - The Kaleidoscope of Bioanalysis, November 16-18, 2016
Barcelona, Spain
Recommendations presented here were influenced by our collaborations on these AAPS white-papers:

**Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides - Harmonized Terminology and Tactical Recommendations, AAPS Journal, 2014**

*The quintessence of immunogenicity reporting for biotherapeutics, Nature Biotechnology, 2015*

**Authors:**
- Gopi Shankar, Ph.D. (*Janssen R&D / Johnson & Johnson*)
- Arno Kromminga, Ph.D. (*European Immunogenicity Platform*)
- Steven Arkin, MD (*Pfizer*)
- Meena Subramanyam, Ph.D. (*Biogen-Idec*)
- Viswanath Devanarayan, Ph.D. (*AbbVie*)
- Laurent Cocea, MD, Ph.D. (*Health Canada, CERB/Clinical Evaluation Division*)
- Daniela Verthelyi, MD, Ph.D. (*FDA, CDER/Office of Therapeutic Proteins*)
- Susan Kirshner, Ph.D. (*FDA, CDER/Office of Therapeutic Proteins*)
- Mark Borgini, MD (*FDA, CBER/Office of Cellular and Gene Therapies*)
- Sarah Yim, MD (*FDA, Division of Pulmonary, Allergy and Rheumatology Products*)
- Christian Schneider, MD (*EMA, Danish National Health and Medicines Authority*)
Clinical ADA data evaluation

Determine sample ADA status

↓

Determine Subject ADA status based on sample ADA results

↓

Describe characteristics of the ADA immune response:
*pre-existing antibodies, ADA Incidence, Titer, NAb, Kinetics, any endogenous cross-reactivities*

↓

Descriptive illustrations

↓

Descriptive/Exploratory Statistics

↓

Determination of clinically impactful ADA threshold of ADA attributes
Terms & Definitions

Simple terminology requiring clarification:

• ADA, Binding ADA, Neutralizing ADA, Non-neutralizing ADA, sustaining ADA, clearing ADA, HAMA, HACA, HAHA, Titer, etc.

Terms used to describe ADA status of a sample:

• **ADA Positive Sample**: when ADA is detected in a sample, the sample is considered positive

• **ADA Negative Sample**: when ADA is not detected in a sample, the sample is considered negative

• **ADA Inconclusive Sample**: when ADA is not detected in a sample but drug is present in the same sample at a level that can produce interference in the ADA detection method, then the negative ADA result cannot be incontrovertibly confirmed and the sample classification for ADA status should be considered inconclusive.

• **Unevaluable Sample**: when a sample could not be tested for ADA due sample loss, mishandling, or errors in sample collection, processing, storage, etc.
Terms & Definitions (contd.)

**Baseline ADA (pre-existing antibodies):** refers to antibodies reactive with the biologic drug molecule before initiation of treatment.

**Treatment-induced ADA:** ADA developed *de novo* (seroconversion) following biologic drug administration *in a subject without pre-existing ADA*.

**Treatment-boosted ADA:** Pre-existing ADA that were boosted to a higher level following biologic drug administration (i.e., any time after the initial drug administration the ADA titer is greater than the baseline titer by a scientifically reasonable margin.

**Terms used to describe ADA status of a Subject:**

- **ADA Positive Subject:** Subject with at least 1 treatment-induced or treatment-boosted ADA positive sample at any time during the treatment or follow-up observation period.

- **Baseline ADA positive Subject:** An ADA positive subject with baseline positive sample(s), regardless of boosting after biologic drug administration.
Sampling Recommendations

ADA should be evaluated in **all study patients** and not only in a symptom-driven manner

*Always* collect a **baseline** sample

**Sampling frequency** during treatment should be designed to maximize the opportunity of detecting treatment-induced ADA and, when applicable, to understand the **kinetics**.

*At least one sample should be collected* following an appropriate period of time **after the last drug administration**.

For the registration (BLA/MAA) of chronic treatments, regulatory authorities typically expect immunogenicity **data through the first year**. Depending on clinical study length, collect samples for testing at: **2 weeks (optional), 1 mo, 2mo, 3mo, 6mo, 9mo, 12mo, 18mo, 24mo, and every year thereafter during treatment**, **and at least one more sample after the last drug administration** when off-treatment persistence of ADA needs evaluation.
Pre-existing Antibodies

ADA at baseline ("prevalence") may have clinical consequences

Report separately to enable analysis of association with clinical outcomes after first dose

Assess treatment-boosting of the pre-existing antibodies to understand association with clinical outcomes after repeat dosing

“boosted by a scientifically reasonable margin”

• Based on a reasonable titer increase; not log-fold...

• Titration with 2-3 fold serial dilution scheme is encouraged; greater dilution is not a good idea.

• titer increase such as a 4-fold (for 2x dilution scheme) or 9-fold (for 3x dilution scheme) is a reasonable margin for change in titer

• Optionally, but certainly for higher risk biologics, a 2-fold increase is considered significant.
Data Analysis & Presentation

Describe characteristics of the ADA immune response: *pre-existing antibodies, ADA Incidence, Titer, NAb, Kinetics, any endogenous cross-reactivities*

- Descriptive illustrations
- Descriptive/Exploratory Statistics
- Determination of clinically impactful ADA threshold/cut-off of ADA attributes
Descriptive characteristics of an ADA immune response

**ADA Prevalence (baseline / pre-existing):**

- Baseline ADA positive subjects as a percentage of the total number of subjects whose baseline samples were tested for ADA.
- Titer range (median, IQR) of the baseline ADA positive samples

**ADA Incidence:**

- **Overall ADA incidence:** combined results of treatment-boosted ADA positive subjects and treatment-induced ADA positive subjects. Compute as a percentage of the total number of evaluable subjects, excluding baseline positive subjects without any samples available after drug administration.
- **Treatment-induced ADA incidence:** computed as a percentage of the total number of evaluable subjects that were ADA negative at baseline. Also report peak positive titer and range (median, IQR) for this group of subjects.
- **Treatment-boosted ADA incidence:** computed as a percentage of the total number of evaluable subjects that were ADA positive at baseline. Also compute the fold-increase in titer (ratio of peak post-administration titer to baseline titer) and range of titer increases (median, IQR).
Descriptive characteristics of an ADA immune response

**ADA Kinetics:**

- **Onset:** the time period between the initial administration of the biologic drug (in a study) and the first instance of treatment-induced ADA.
  - Compute the “median time to ADA development” and the inter-quartile values Q1 and Q3, which can enable an understanding of onset of ADA in half, 25%, and 75% of the treated subjects, respectively.

- **Duration:** refers to the longevity of treatment-induced ADA. Criteria for “Transient” vs. “Persistent” should be based on the natural clearance of endogenous human Ig.
  - If ADA are predominantly IgG1/2/4, ADA response lasting > 16 weeks can be considered as Persistent. This is because t1/2 is 21-25 days, So 97% of the IgG would have been eliminated after 5 half-lives (16 weeks).
  - If ADA are predominantly IgM, IgA or IgG3, use 5 weeks.
  - Do not based this on when sampling was stopped (i.e., “transient” because seronegativity was achieved during study is wrong)
  - “Transient” vs. “Persistent” describe the longevity of the ADA. It does not imply negative or positive association with clinical consequences!
Descriptive characteristics of an ADA immune response

**Transient ADA:**
- Treatment-induced ADA detected only at one sampling time point during the treatment or follow-up observation period (excluding the last sampling time point, as a conservative measure)

or

- Treatment-induced ADA detected at 2 or more sampling time points during the treatment (including follow-up period, if any), where the first and last ADA positive samples (irrespective of any negative samples in-between) are separated by a period less than 16 weeks, and the subject’s last sampling time point is ADA negative.
Persistent ADA:

• Treatment-induced ADA detected at 2 or more sampling time points during the treatment (including follow-up period, if any), where the first and last ADA positive samples (irrespective of any negative samples in between) are separated by >= 16 weeks

or

• Treatment-induced ADA incidence only in the last sampling time point of the treatment study period, or at a sampling time point with less than 16 weeks before an ADA negative last sample. [By conservative inference]

• Although rare, when a patient population in a study is predominantly found to develop IgG3 or IgA, a 5 week period should be applied to modify the definitions of transient and persistent ADA (instead of the 16 weeks). This is because IgG3 and IgA have shorter half-lives than other IgGs (IgG3, 7 days; IgM and IgA, 5 days).
Descriptive characteristics of an ADA immune response

**NAb Incidence and kinetics:** when study results indicate distinct NAb-containing versus non-Nab containing subject groups, it is useful to look at NAb incidence and kinetics separately for each group in the same manner as described above for ADA.

**Cross-reactivity:** when a biologic drug molecule is identical or nearly identical to an endogenous protein (whole or in part), it is important that the *cross-reactivity of the ADA with the endogenous protein* be evaluated because of the potential to cause an autoimmune-like syndrome.

- Compare the kinetics and titers of the cross-reactive ADA versus the whole-drug ADA, and relate this to the worsening of disease.
Treatment-induced ADA incidence & kinetics
This plot includes ADA positive patients for whom the ADA onset time is at least 120 days before the last visit OR they were ADA negative by the last visit. Reference lines represent the quartiles.
Cumulative treatment-induced ADA incidence & kinetics

After aggregation of data, cumulative profile can be shown as below:
Descriptive illustration of ADA Titer Kinetics

ADA Incidence:

*Numerator alphabet*: # ADA positive subjects at that sampling time point

*Denominator alphabet*: # total subjects evaluable for immunogenicity

(Percentage): proportion of ADA positive subjects

*outliers
Descriptive illustration of ADA Titer Kinetics
Illustration with example clinical data
Descriptive statistics for ADA Kinetics

Comments

This approach obviates definitions for ADA immune response kinetics – the onset (early/late) or duration (transient/persistent).

*These evaluations are hard to interpret for very small sample size and/or incidence rate.*

*Complex studies with multiple arms and satellite studies may not allow for statistical assessments.*

*Applies only to Treatment-Induced ADA* (treatment-boosted ADA are different mechanistically).

**ADA Onset:** report the quartiles (Q2 is “Median time to antibody formation”)

- **Interpretation:** “When half of the ADA + subjects seroconverted” OR “when 75% (majority) of the ADA+ subjects seroconverted”

**ADA Duration:** report the quartiles (Q2 is “Median time of antibody duration”)

- **Interpretation:** “The ADAs lasted Q2 months in half of the subjects and Q3 months in 75% of the subjects”
Assessment of ADA vs. clinical outcomes

Requires an **integrated analysis** of PK (serum conc., clearance rate), PD (when applicable), Efficacy, and ADRs (acute and non-acute), in relation to the intended dosage (tested in pivotal trials).

For determining a clinically relevant “threshold”,
• ROC and Tree-based analysis of the ADA characteristics vs. PK/PD, changes in disease parameters (efficacy), and levels of ADR, may be used.
  ▪ **Not feasible/relevant for small sample size and/or incidence rate.**
Influence of ADA on Clinical Efficacy

% of Patients with desired efficacy vs Time (weeks/months)
Median longitudinal efficacy profiles for ADA positive & negative patients are plotted separately. Error bars represent standard error of median (estimated using MAD). Efficacy is reduced in ADA positive patients at several time points.
ADA vs. Clinical Efficacy
Illustration with example clinical data (Titer, Duration, Onset)

Relationship between the 52 week efficacy (% reduction in disease activity score) vs. maximum titer during the 52-week period, ADA duration and ADA onset time are summarized here. Each point in the graph is a unique ADA positive patient. Patients with ADA onset later than 300 days were excluded. Trend is approximated using smoothing spline. Efficacy trends lower with longer ADA duration ($p<0.05$), and also with higher titer with a breakpoint at around 120.
Statistical evaluation of ADA vs. Clinical endpoints

Some important considerations

Trends are usually non-linear, often with a step-wise association.

Therefore, linear-based models such as Logistic or Linear Regression may not capture the associations well.

Need flexible nonparametric and/or nonlinear models to evaluate these associations.

The ADA parameters (onset, titer, duration) are typically not independent. Need to model their interactions.

They may correlate in a collective manner (multivariate).

What statistical approach addresses these needs?
Statistical evaluation of ADA vs. Clinical endpoints
Some considerations (contd.)

For analyzing the *impact of each ADA parameter independently on Clinical endpoints (univariate)*, ROC analysis or univariate Tree model is useful.

- Nonparametric & nonlinear
- Readily provides a threshold estimate that optimizes the association (e.g., sensitivity/specificity).

“*Generalized Additive Models*” (GAM) may be useful for visualizing and modeling the collective association of ADA parameters on clinical data.

For characterizing the *collective impact of ADA parameters on clinical endpoints*, Multivariate Tree-based model may be useful.

- Accounts for interaction between ADA parameters, in addition to modeling the nonlinearity and being nonparametric.
- Output: “Decision Thresholds on multiple ADA parameters”
Each ADA parameter vs. Clinical outcome
Illustration of ROC analysis (univariate)

Titer threshold at 120 provides ~ 76% Specificity & ~74% Sensitivity.

i.e., 74% of patients with favorable efficacy have Titer < 120, and 76% of patients with poor efficacy have Titer > 120.

Increasing the Titer threshold to 240, results in ~68% Specificity and ~ 83% Sensitivity.

This analysis permits the association of only one feature at a time.
Need Multivariate Tree-based model for assessing joint effect of multiple ADA parameters.
**ADA vs. Clinical Safety (AE incidence)**

**Visualizing & Modeling (Generalized Additive Models)**

When ADA duration is < 4 months (16 weeks) and Titer < 15, there is no impact on Clinical AE. Note the sharp correlation to clinical AE with the steep increase in the slope/surface when ADA duration > 16 weeks & Titer > 15.

*Color shades denote probability of clinical AE. Probability increases as the color changes from dark to light blue.*
ADA vs. Clinical Safety (AE incidence)
Illustration of Tree-based Model (multivariate)

Higher AE incidence (30%) for patients with ADA titer > 20 and onset within 6 months.

18% of patients have AE (n=16)

*Significance of onset-time is not strong.

No AE incidence for patients with low ADA titer (<20)
ADA vs. Clinical Efficacy

Illustration of Tree-based Model

Efficacy drops to 37% for patients with ADA duration > 442 days (~15 months)

Efficacy stronger (80%) for patients with ADA duration < 16 weeks.

*LogWorth > 1 implies p < 0.01 (after multiplicity adjustment),
ADA & NAb vs. Clinical Efficacy (Role of NAb?)

Illustration of Tree-based Model (multivariate)

Efficacy drops to 37% for patients with ADA duration > 442 days (~15 months)

69.4% efficacy (n=61 patients)

Efficacy drops to 71% for NAb+ patients

Efficacy stronger (80%) for patients with ADA duration < 16 weeks.

NAb Negative
Mean = 86, n=19

NAb Positive
Mean = 71, n=12
Limitations with ROC/Tree-based analysis, and proposed improvements

Not very stable.

- Significance of ADA parameters may vary greatly due to minor changes in the data.

ADA Threshold estimates may be highly noisy/variable, especially for smaller sample size or incidence (typical).

To obtain more robust threshold estimates and evaluate variability (confidence interval), we now propose methods to extend the univariate ROC and multivariate Tree analysis via a bootstrap resampling scheme.

- (next two slides)
For increasing robustness of univariate thresholds from ROC or Tree analysis, we propose this resampling method:

**BATTing**: Bootstrapping & Aggregating Thresholds from ROC/Trees

Devanarayan, 1999

---

Threshold estimated from this approach is robust to small perturbations in data, outliers, etc.

Robust Threshold = Median of this distribution

Spread of this distribution reflects the variability.
For more robust multivariate thresholds (signature), we propose this “sequential resampling” extension.

**Sig+: subgroup of patients with lowered efficacy/PK or AE**

**Model Growing within the potential Sig+ group**
- Get the BATTing threshold for each *unused* ADA parameter.
- The best predictor is selected to split the current sig+ group.
- This procedure continues in the new Sig+ group.

**Stopping Rule:**
- The new added predictor goes through the likelihood ratio test for significance.
Additional comments on the statistical evaluations of ADA parameters vs. Clinical PK/Efficacy/Safety

As evident from the methods and illustrations, *this analysis should ideally be performed by experienced biostatisticians.*

Inadequate analyses may result in *false negative and false positive claims on ADA association to clinical outcomes!*

**False-Negative Associations:**

- Due to inherent non-linearity, *linear-based models are inadequate.* May lead to incorrect conclusions on lack of or no association.
- Need flexible nonparametric & nonlinear methods that account for stepwise nonlinear trends & interactions between ADA features (Tree, GAM, etc.).
- Need resampling-based methods to robustify results & evaluate variability.

**False-Positive Associations:**

- Beware of over-fitting. Performance claims should not be based on the same data that were used to derive the associations & thresholds.
- Need rigorous internal cross-validation to reduce bias in the estimates of model accuracy and performance. Follow-up with external validation.
Summary: Clinical ADA data evaluation

Determine sample ADA status

Determine Subject ADA status based on sample ADA results

Describe characteristics of the ADA immune response:
pre-existing antibodies, ADA Incidence, Titer, NAb, Kinetics, any endogenous cross-reactivities

Descriptive illustrations

Descriptive/Exploratory Statistics

Determination of clinically impactful ADA threshold of ADA attributes