PharmaSUG China 2025 - Paper DV-147

Immunogenicity Data Visualization in Vaccine Development: A Case Study with Data Generated by ImmunoDummyGen

Bingqin Liu, Mengran Ma, Pfizer (China) Research and Development Co.,Ltd. China

ABSTRACT

Immunogenicity analysis is a crucial component of vaccine clinical trials, providing insights into a vaccine's ability to elicit immune responses and guiding decisions on efficacy, safety, and dosing. However, challenges such as the sensitivity of early-stage immunogenicity data and the time-consuming nature of developing analytical programs can hinder efficient analysis. To address these challenges, we present ImmunoDummyGen, a Python-based tool that integrates with SAS macros to generate synthetic, privacy-compliant immunogenicity data sets. With a user-friendly graphical interface (GUI), ImmunoDummyGen enables the simulation of trial-specific dummy data by incorporating metadata such as visit schedules, biospecimen collection timepoints, and assay plans, facilitating the preparation of statistical analyses.

This paper demonstrates the use of ImmunoDummyGen-generated dummy data and introduces a comprehensive visualization framework for immunogenicity analysis in vaccine development. We illustrate a range of visualizations - such as antibody titer bar plots, longitudinal response curves, and spaghetti plots of ratio data - that simplify the interpretation of complex immunological data. These visualizations enhance communication across various stakeholders and support the effective presentation of vaccine performance. By bridging the gap between protocol design and data constraints, this solution reduces programming time, mitigates risks associated with tight timelines, and improves operational efficiency and regulatory compliance in vaccine development.

INTRODUCTION

Immunity to a disease is conferred by the presence of specific antibodies in the body, which can be acquired either actively or passively. Unlike natural immunity, which arises from direct exposure to the disease pathogen through infection, vaccine-induced immunity involves the administration of a killed or attenuated form of the pathogen. As a result, vaccine development has become one of the most prominent areas of research in recent years. Given the underlying rationale for how vaccines protect individuals, immunogenicity analysis is a critical component of vaccine clinical trials. It not only helps assess the safety of the vaccine but can also serve as a surrogate endpoint for efficacy in many cases. However, due to the sensitive and time-consuming nature of immunogenicity testing, results are typically unavailable until the end of study. Therefore, dummy data are often employed during the trial process to facilitate the work of SAS programmers.

This paper introduces a Python-based tool that integrates with SAS macros to generate dummy data for a case study. Using the data from case study, we demonstrate various visualizations of immunogenicity results and illustrate how these results would be interpreted throughout the study.

DATA SOURCE

Immunogenicity analyses are often performed with collecting blood draws. The blood samples in any study will be used to measure parameters that will address the objectives of this study. Blood samples will be collected from selected participants at selected sites only strictly following documented standard procedures. In addition, blood will also be collected at visits specified in documents and sent to the site's designated processing laboratory for further processing. Common sources for assays are listed as following in Table 1. Note that both the sample collection process and the process of transferring to be assessed are highly cautious and confidential. So due to the sensitive and time-consuming nature of immunogenicity testing, results are typically unavailable until the end of study. Therefore, dummy data are often employed during the trial process to facilitate the work of SAS programmers.

Biospecimen Type	Assay

Serum	Antibody
Peripheral Blood Mononuclear Cells (PBMC)	T-cell response, B-cell response
Whole Blood	HLA typing

Table 1. Common source for different assays

DATA SIMULATION

ImmunoDummyGen has a user-friendly graphical interface (GUI) built by Python Streamlit package and the background Python algorithm integrates with SAS macros to generate synthetic, privacy-compliant immunogenicity data sets. Users utilize this tool by inputting both biospecimen plan and study directory where the raw data sets/SDTMs store at. There are both biospecimen plan loading button and entry textbox for study directory present on the Python Streamlit webpage. Biospecimen plan contains not only Biospecimen Schedule & Types (Table 2) including visits and number of participants at each visit with their corresponding treatment groups (Table 3) but also sample aliquots and their associated assays / assays for biospecimen type (Table 4). ImmunoDummyGen then calls SAS utility code to read, manipulate the biospecimen plan, restructure all the information inside biospecimen plan and finally return a consolidated excel spreadsheet table. The conjugated/extracted table returned by this tool is a combination of participant treatment or subset group descriptions, participant groups, number of participants per group, assay list, biospecimen type, visit numbers, visit names which all those assays are collected at, etc. This generated excel file will be present on GUI webpage shown as a small size tibble for users to visualize. Users can download this excel file to further customize the contents by deleting redundant assays which won't be used and adding more information such as manually assigning LLOQ values and participants' treatment groups if needed. Then after modification, users save all the updates by clicking the "Save Update" button on webpage. The background server will use this updated version of excel to continue the data simulation process.

VISIT ID	VISIT NAME	Biospecimen Type	Projected Total no. Participants
10001	V1_VAX1	SERUM	100
10002	V2_WEEK1	SERUM	100
10003	V3_MONTH1	SERUM	100
10004	V4_MONTH3	SERUM	100
10005	V5_MONTH6	SERUM	100
10001	V1_VAX1	PBMC, WHOLE_BLOOD	70
10002	V2_WEEK1	PBMC	70
10003	V3_MONTH1	PBMC	70
10004	V4_MONTH3	PBMC	70
10005	V5_MONTH6	PBMC	70

Table 2. Sample BIOSPECIMEN PLAN: BIOSPECIMEN SCHEDULE & TYPES Table

Ref no.	Participant Treatment or Subset Group Descriptions *(as applicable)*	No. of Participants per group	Participant Groups (LIMS)
1	Vaccination A	50	DEFAULT
2	Vaccination B	50	
3	Subset for PBMC	70	PBMC_Subset_List

Table 3. Sample BIOSPECIMEN PLAN: Participant Groups Table

BIOSPECIMEN TYPE: SERUM/PBMC/ WHOLE_BLOOD						
Priority	**AP#**	**Lab**	**Assay Name**	**Analysis(es)**	**Visit Name(s)**	**Participant Group (LIMS)**
1	N/A*	VRD	COV2 MNT 384w	COV2_WT_NT50 COV2_OC_NT50 COV2_BA2_NT50	V1_VAX1 V3_MONTH1 V4_MONTH3 V5_MONTH6	DEFAULT

Table 4. Sample BIOSPECIMEN PLAN: Priority of sample aliquots and their associated assays / assays for biospecimen type

Based on the extracted excel table containing assays information generated by Python, dummy SAS code will be called to start the process of simulating both raw data and IS SDTM (Figure 1). Firstly, SAS code will access the extracted table and merge it with electronic sample tracking raw data which contains each participant's visit(s) and corresponding Date/Time of specimen collection, subcategory for biospecimen event, barcode, etc. Secondly, by merging with the standard VARD assay map list, we could easily get assay description, Unit of Measure (UoM), specimen type for dummy raw data. In addition, VARD assay map list also provides standard wording for ISMETHOD, ISSPID, ISTEST, ISTESTCD and ISCAT in SDTM.

After basic data structure is ready, the next step is to generate results for each assay. By applying for RANNOR Function in SAS which returns a random variation from a normal distribution, we could simply generate random numbers with consideration of LLOQ for serum assays in numeric results. Except for numeric results, character results can be included based on different assays requested by biospecimen plan. Results for HLA-Typing assays are expressions of alleles. Therefore, we randomly select results by linking HLA alleles with alleles result list that download from **Immuno Polymorphism Database** to generate dummy results. There are also categorical results, for example, results for CD4+ T-cell response are collected as positive response, no response, and not evaluable, the categorical would be randomly generated by RAND function in SAS and convert numeric outcomes into character results.

With the simulation of all the above info together, we could utilize the simulated CSV raw data and IS SDTM to help us further generate ADAM data sets and any type of TLF as needed.

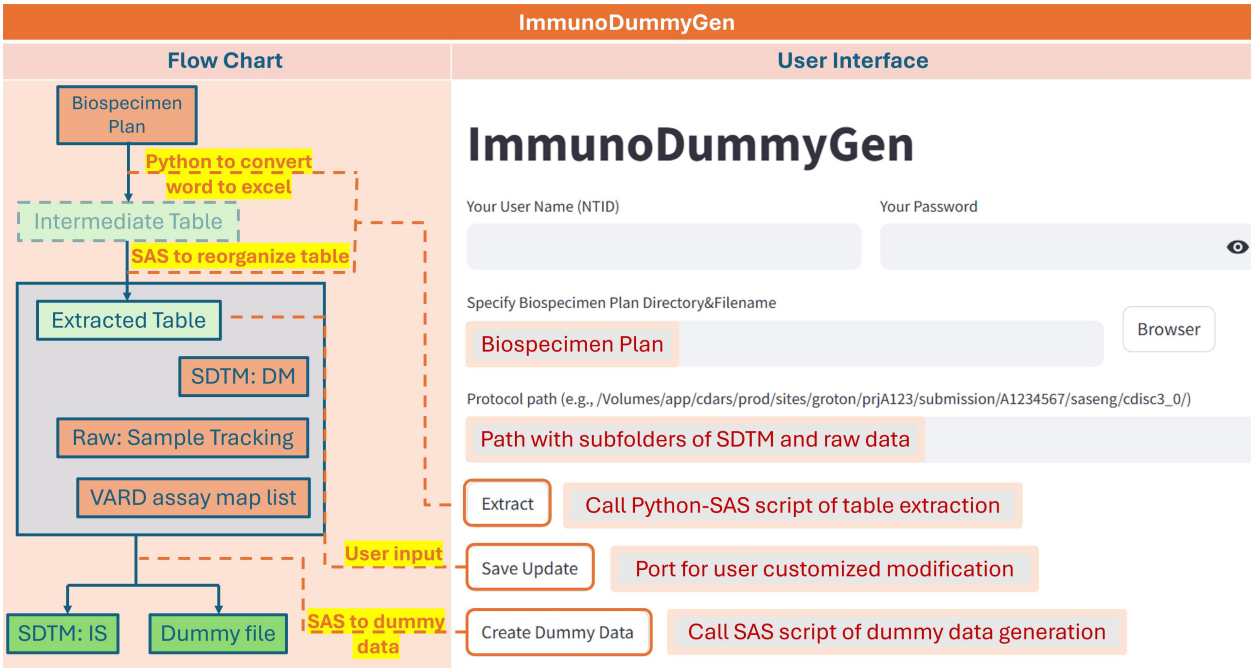

Figure 1. ImmunoDummyGen Working Flow

IMMUNOGENCITIY ANALYSIS

RCDC

The utility and advantages of Reverse Cumulative Distribution Curves (RCDCs) in summarizing immune response profiles in vaccine studies have been demonstrated by numerous researchers. RCDCs are not only straightforward to interpret but also provide sufficient information for statisticians to conduct further statistical analyses. Figure 2 presents an RCDC generated using synthetic data from the case study. The graph clearly illustrates that Vaccination A induced a greater increase in the relevant antibody titer, making it a more favorable option for participants.

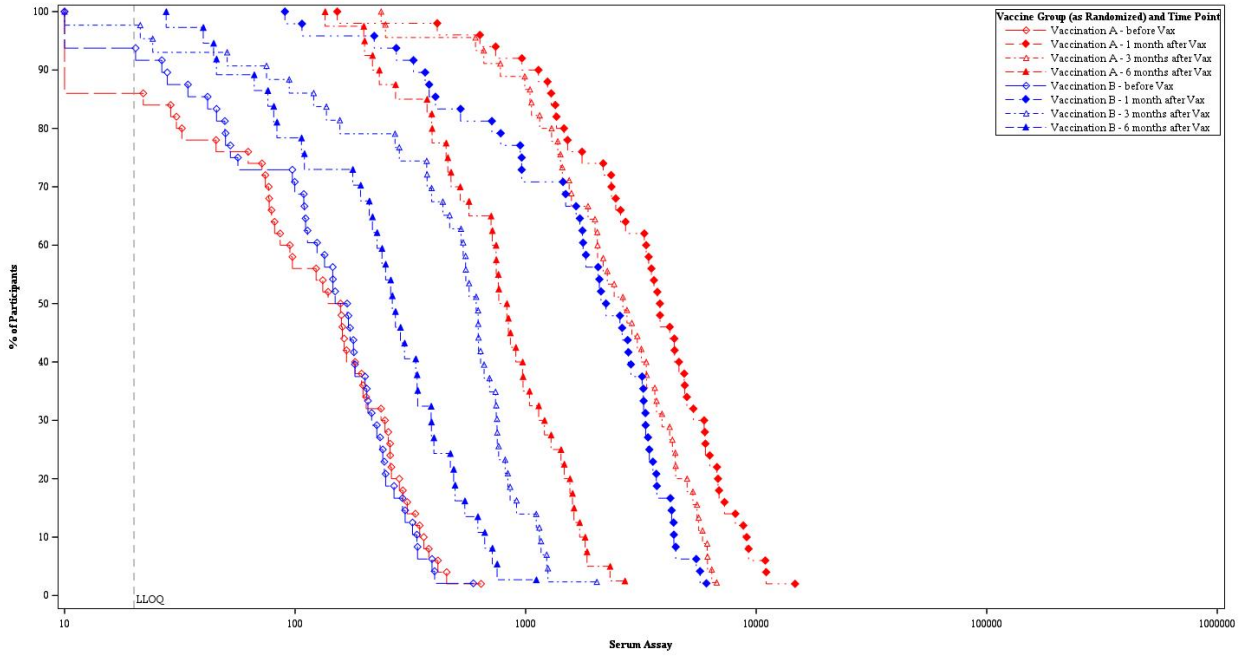

Figure 2. Reverse Cumulative Distribution Curve for Serum Assay by Vaccine and Visit

BAR CHART

Bar charts are among the most widely used visualizations in the pharmaceutical industry, as they effectively represent various scenarios related to the progression of antibody. Figure 3 is a typical plot used in vaccine studies to illustrate trends following vaccination, comparing geometric mean in different vaccines as well as variations within each vaccine group. Scatter points and confidence intervals are also included in the plot to display the distribution across each vaccination group, subgroup, and visit. While Vaccination A demonstrates superior performance compared to Vaccination B, no significant difference is observed between subgroups 1 and 2 when considering baseline status.

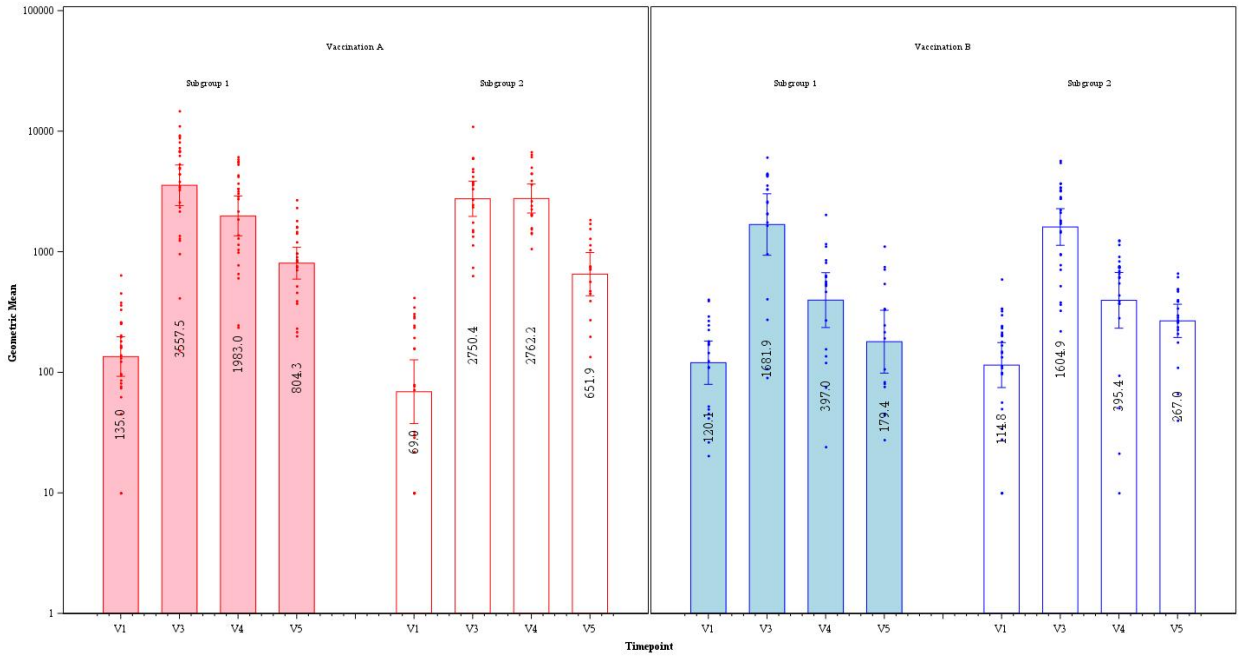

Figure 3. Bar Chart for Continuous Serum Assay by Baseline Status

In addition to continuous serum assays, which can be presented using bar charts, categorical PBMC assay results can also be illustrated effectively. The PBMC assay results in Figure 4 are categorized into three groups: positive response, no response, and not evaluable. A bar chart is an ideal visualization for this type of data, as it clearly shows the distribution across each vaccine, subgroup, and visit. Both subgroups of Vaccination A exhibit slight changes in the number of positive responses over time.

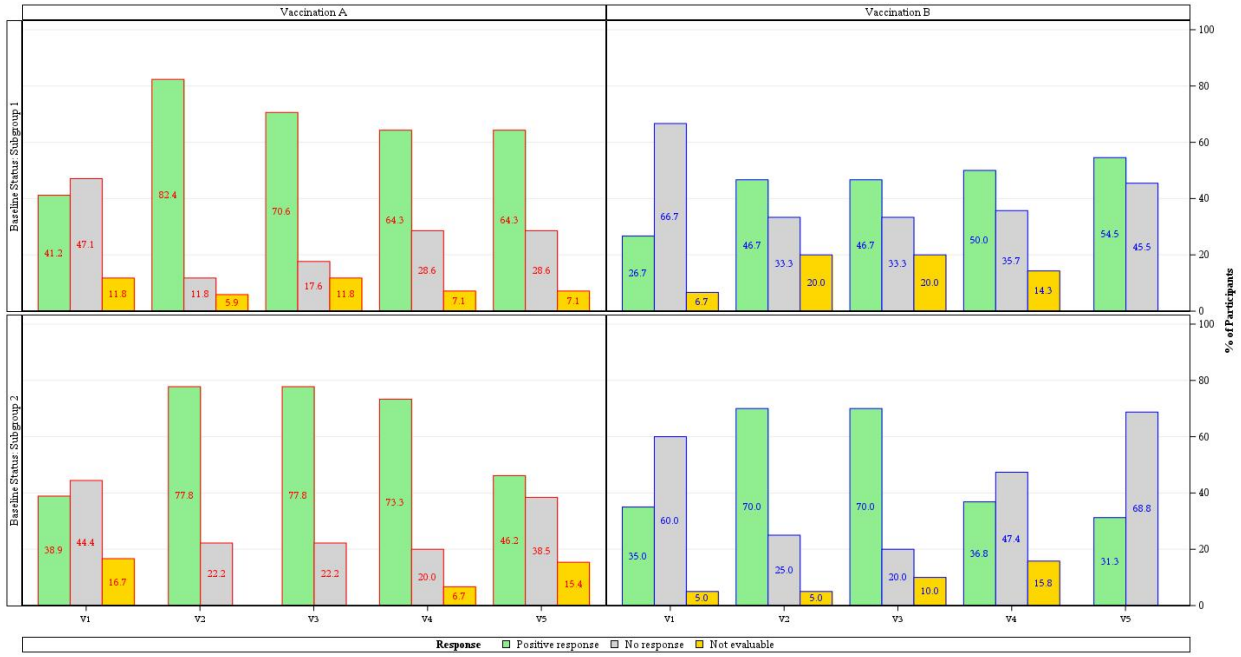

Figure 4. Bar Chart for Categorical PBMC Assays

SERIES PLOT

While trends in geometric mean can be explored using bar charts, the relationship between assay concentrations is more clearly depicted in a series plot. Series plots are particularly useful for determining the effective duration of a vaccine's protection and can help assess the need and timing of booster doses. With advancements in medical science, the mechanisms underlying many diseases and the levels of antibodies required for effective protection are increasingly well understood, such as the 0.35 µg/ml threshold for pneumococcal vaccines. By identifying the threshold for serum assays in Figure 5, it becomes possible to infer that Vaccination A provides protection for up to 6 months, while Vaccination B offers protection for only more than 3 months.

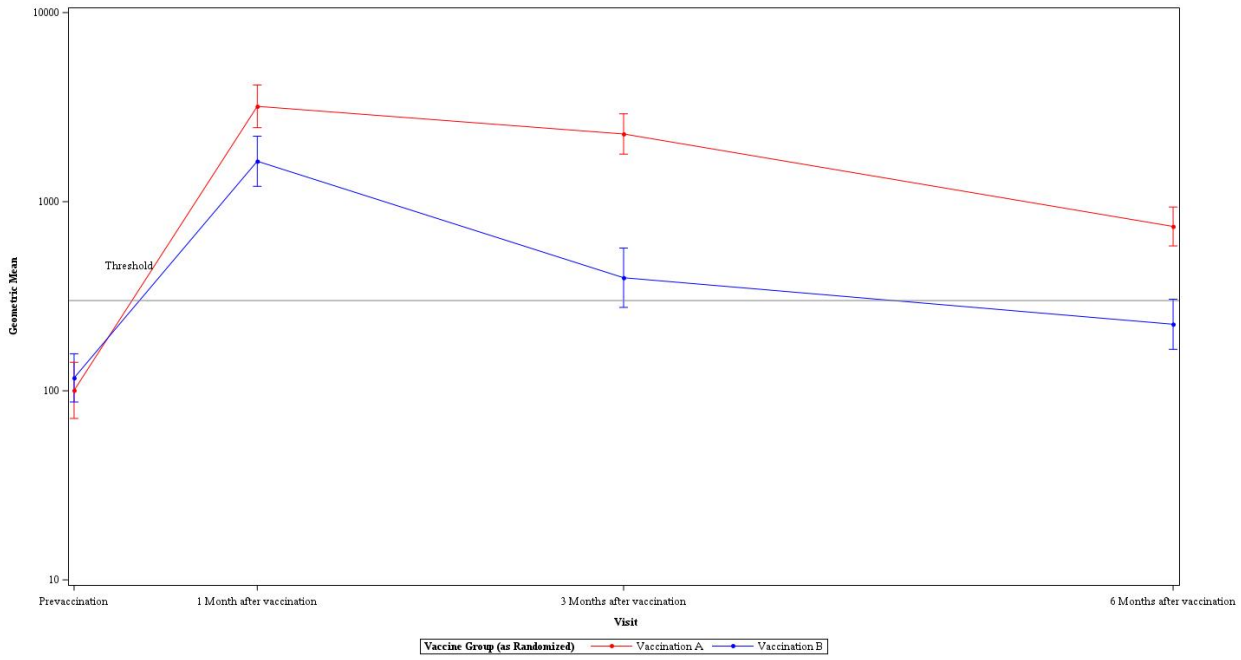

Figure 5. Series Plot for Serum Assay Against Time

More complex patterns in series plots are also employed in immunogenicity analysis for assays. One example is Figure 6, where the data for all participants are displayed on a single page. These trends should align with the overall geometric mean series plot for each group, while also providing additional information on the individual within the plot. Similarly, Figure 7 illustrates the trend in the ratio between two PBMC assays, offering a more precise assessment of the vaccine's efficacy. Figure 8 presents individual series plots for the ratio of PBMC assays, with different assays displayed together and connected by a line. The slope of this connecting line reflects the ratio between the PBMC assays. In contrast to Figure 7, Figure 8 provides information on both individual assays and their ratios.

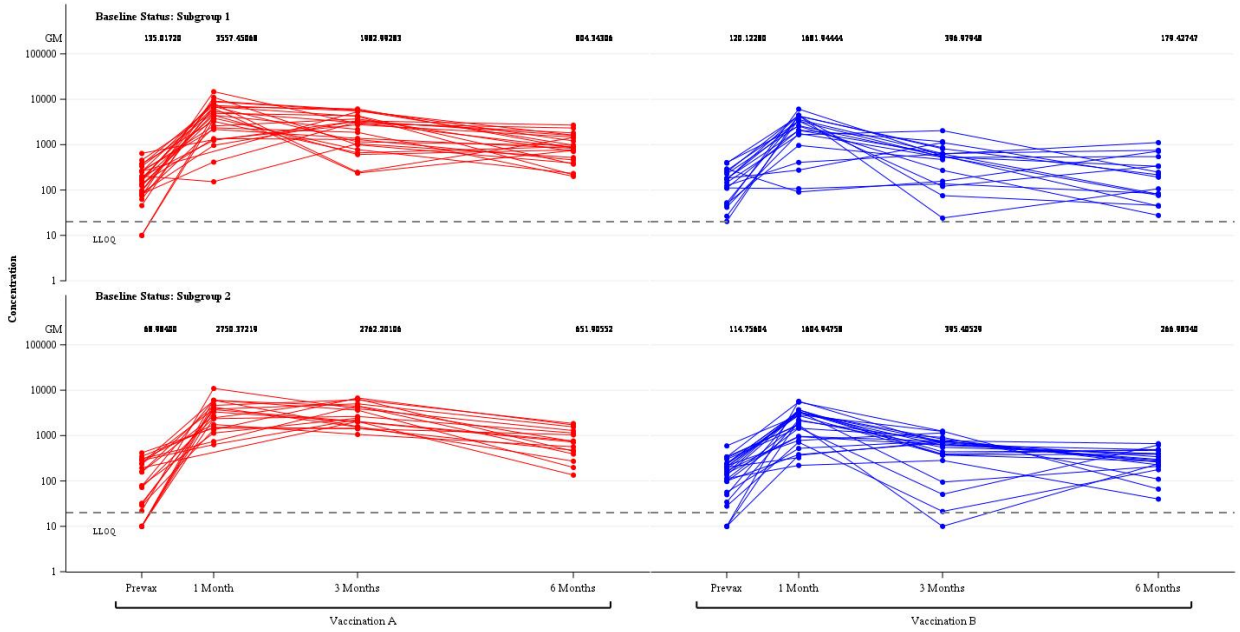

Figure 6. Individual Series Plot for Serum Assay Against Time

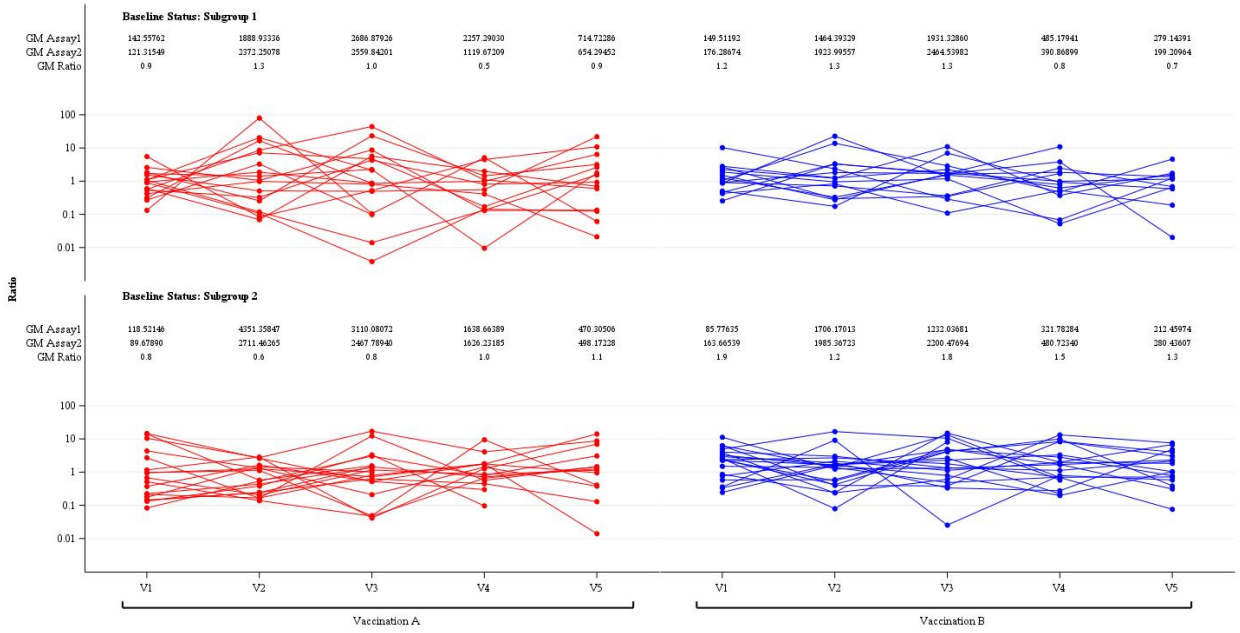

Figure 7. Individual Series Plot for Ratio of PBMC Assays Against Time

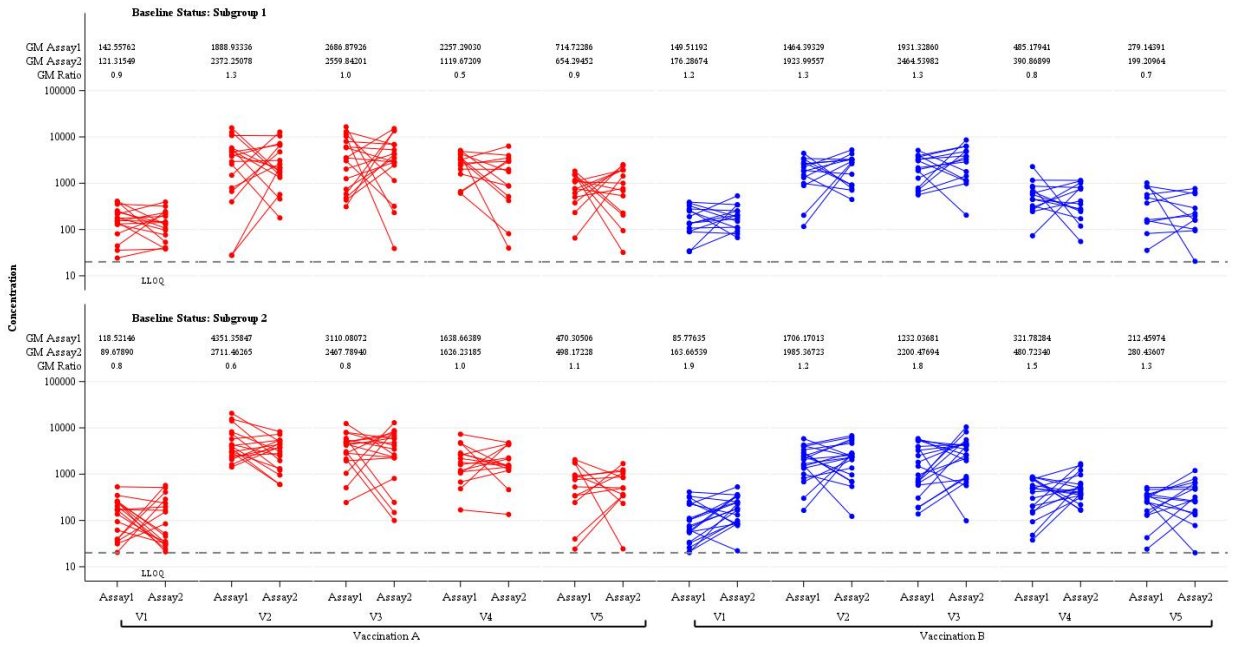

Figure 8. Individual Series Plot for Two PBMC Assays Against Time

SCATTER PLOT

Scatter plots are commonly used as components within other visualizations, such as the scatter points in Figure 3, or in conjunction with series plots. However, when examining correlation, scatter plots are the most appropriate choice. Figure 9 presents a scatter plot that illustrates the correlation between related PBMC assays. The 45° reference line aids in clearly identifying the positive relationship between the variables.

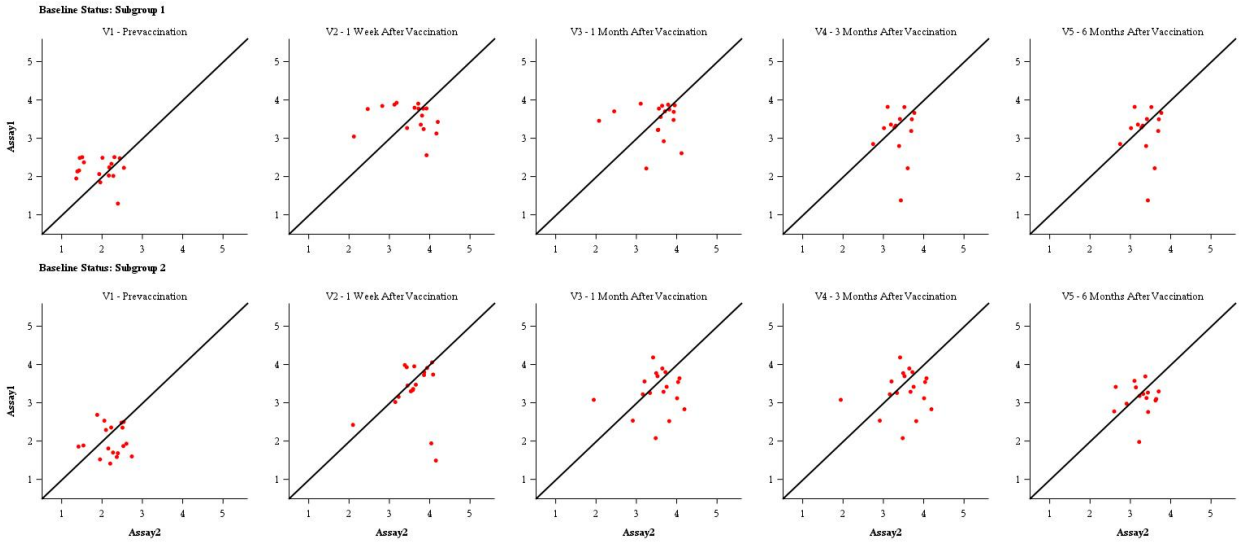

Figure 9. Scatter Plot for Correlation Between Two PBMC Assays

BOX PLOT

Box plots are also highly effective for displaying geometric means. Compared to bar charts, box plots offer a more comprehensive view, including individual data points, mean, median, quartiles, and confidence intervals. In the immunogenicity analysis of PBMC assays, the percentage above baseline

quartiles provide valuable insights for statisticians and clinicians. Therefore, this additional information is included in the plot header. It proves useful, especially when some scatter points overlap, enhancing the interpretability of the data.

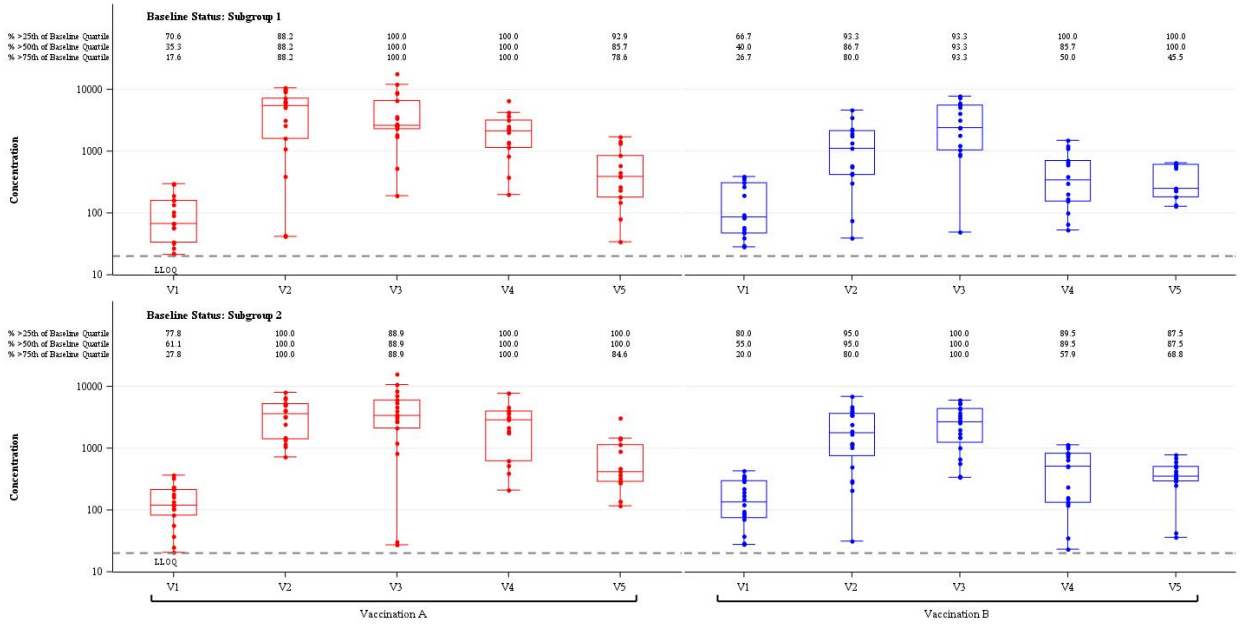

Figure 10. Box Plot for PBMC Assay Concentration

PIE CHART

As noted in Table 1, HLA assay results are also included in the immunogenicity analysis. However, unlike most immunogenicity assays, HLA assays are presented as text, representing the gene expression of HLA. HLA haplotypes may exhibit either identical or distinct expressions. We recommend using a pie chart to present HLA results, as it effectively illustrates the allelic diversity of HLA genes across different groups. The frequency distribution of the alleles, as shown in Figure 11, is represented by the area of the segments in the pie chart. This visualization facilitates the exploration of potential biases in the immune response to the vaccine across different populations.

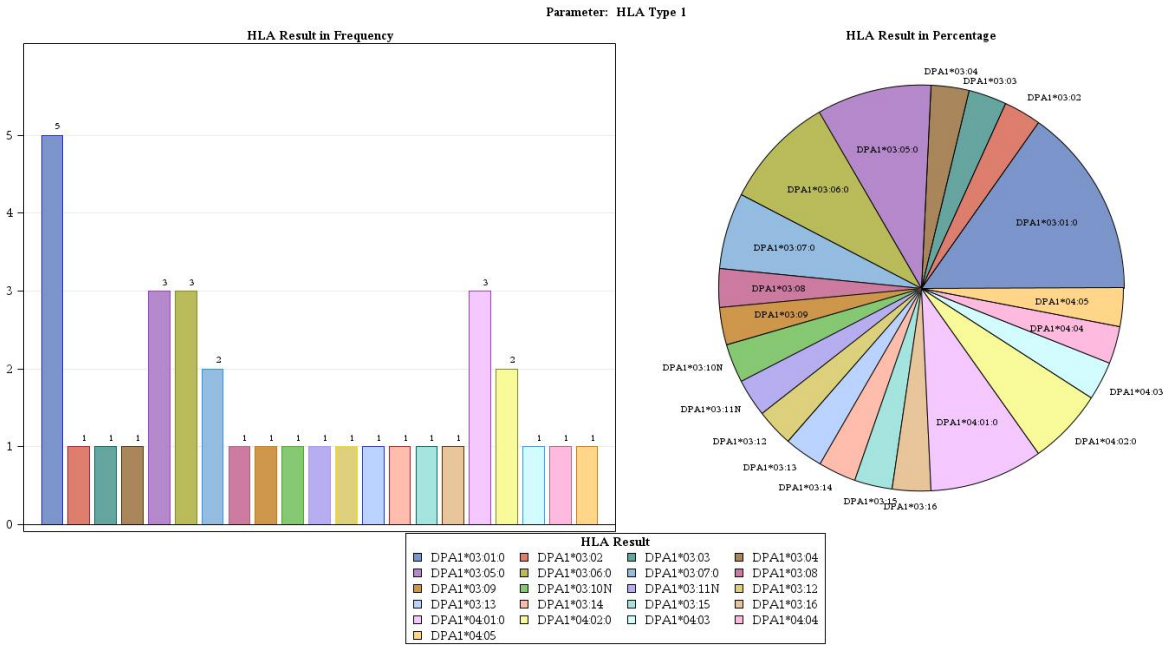

Figure 11. Pie Chart for HLA Haplotype Distribution

CONCLUSION

In this paper, we presented ImmunoDummyGen, a Python-based tool designed to generate synthetic immunogenicity data for vaccine development studies. Through its integration with SAS macros, the tool facilitates the generation based on diverse data sets and enables the creation of multiple types of immunogenicity results. As vaccine studies increasingly rely on complex data analysis, tools like ImmunoDummyGen are helpful in simplifying dummy data generation and enhancing the preparation of programming work.

Based on dummy data, this case study demonstrated the utility of these visualizations in evaluating vaccine performance, comparing subgroups, and assessing the need for booster doses. These visualizations are critical for guiding further statistical analysis and decision-making during the vaccine development process.

REFERENCES

Noel Smith. 2021. “Human PBMC-based assays for the immunogenicity risk assessment of therapeutic peptides.” Accessed June 24, 2025. https://www.fda.gov/media/146114/download

U.S. Centers for Disease Control and Prevention. 2024. “Immunity Types.” Accessed June 24, 2025. https://www.cdc.gov/vaccines/basics/immunity-types.html.

World Health Organization. 2013. “Correlates of vaccine-induced protection: methods and implications.” World Health Organization. https://iris.who.int/handle/10665/84288.

RD Small, A Ozol-Godfrey, L Yan. 2019. “On the use of nonparametric tests for comparing immunological Reverse Cumulative distribution curves (RCDCs).” *Vaccine,* 37:6737-6742.

Marshall JS, Warrington R, Watson W, Kim HL. 2018. “An introduction to immunology and immunopathology.” *Allergy Asthma Clin Immunol,* 14(Suppl 2):49.

Choo SY. 2007. “The HLA system: genetics, immunology, clinical testing, and clinical implications.” *Yonsei Med J*, 48(1):11-23.

Pillai NE, Okada Y, Saw W-Y, Ong RT-H, Wang X, Tantoso E, et al. 2014. “Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations.” *Hum Mol Genet*, 23:4443–51.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Bingqin Liu
Enterprise: Pfizer (China) Research and Development Co.,Ltd.
E-mail: Bingqin.liu@pfizer.com
Name: Mengran Ma
Enterprise: Pfizer (China) Research and Development Co.,Ltd.
E-mail: Mengran.ma@pfizer.com