LEM'S IMMUNOGENICITY ANALYSIS NOW – USER GUIDE

Updated 11/25/19

Table of Contents:

1. Introduction

1.1 Purpose

This document is meant to serve as a reference for all users in order to enlighten and gain an understanding of how to explore all avenues of functionality within the tool. It has been developed and deployed with the help and interests of the following groups in mind:

- *Scientific Implementation Group (SIG)*
- *Clinical Advisor Group (CAG)*
- *Clinical Laboratory Operations (CLO)*

This tool is meant to automate several data analytics processes for various vendor cumulative files, study reports, and CLRM files. Standardized formatting is required for this web-based tool to execute functions correctly (Section 2). Logical checks are also built into the tool to flag discrepancies in data entries and identify inconsistent testing between tiered testing results.

1.2 Scope

The application includes generalized calculations for samples and subjects across all visits in each dataset. Standard summary statistics built within this tool include: number of samples tested and detected in each tier, putative positive rate, confirmed positive rate, number of evaluable subjects, baseline positive rate, treatment emergent rate, treatment-induced and treatment-boosted rates, and number of subjects who are unevaluable. This application has been prototyped using cumulative files from the following vendors: BAL, Eurofins, PBI, PPD, and Wuxi.

1.3 System organization

The R programming language and Shiny package are the primary software tools used to run this application.

R version 3.6.0 (2019-04-26), nicknamed "Planting of a Tree"

Other installed packages and their purpose throughout the script:
dplyr_0.8.1 (filtering and transforming data)
DT_0.6 (provides R interface for data tables on web pages)
ggplot2_3.1.1 (creates aesthetic plots)
readxl_1.3.1 (loads tabular data from Excel spreadsheets)
reshape2_1.4.3 (implemented when creating pivot tables)
shiny_1.3.2 (builds interactive web apps with R)
shinyjs_1.0 (JavaScript-like operations to enhance user-experience)

shinythemes_1.1.2 (includes several Bootstrap themes for styling)
shinyWidgets_0.4.8 (custom input controls and user interface components)
xlsx_0.6.1 (provides R functions to read/write/format Excel 2007)

2. Preprocessing Data Files

2.1 Trimming files

Currently, this R Shiny application requires files to be uploaded in the .xlsx format. On load, the application is only searching for a few specific columns to conduct analyses and build calculations. Thus, it is recommended to delete unnecessary columns. This will not only reduce processing power and time to load data, but also boost user experience and allow for a much simpler interface on the page.

2.2 Modifying columns to the standardized format

R is case-sensitive when processing columns by name. In a nutshell, correct capitalization, spacing, and spelling are essential to making this Shiny app execute properly. The order of the columns is not critical to how the application performs. At a minimum, the following four columns (and their exact syntax) are required for the tool to run properly:

Subject

Visit

Tier2

Tier3

Larger datasets with multi-tier testing can also be processed with the correct column names:

Tier1

Tier2b

Tier2c

Tier2d

Tier4

Tier4b

Tier4c

Tier4d

Tier4e

Tier4f

1	Subject	Visit	Tier1	Tier2	Tier3	Tier4
2	100-00102	BL/V2	NOTDETEC			
3	100-00102	V3	NOTDETEC			
4	100-00102	V4	NOTDETEC			
5	100-00102	V6	NOTDETEC			
6	100-00102	V7	NOTDETEC			
7	100-00102	EV1	NOTDETEC			
8	100-00102	EV2	DNR	DETECTED	1:10	NOTDETEC
9	100-00102	EV3	DNR	DETECTED	1:10	NOTDETEC
10	100-00102	EV6	NOTDETEC			
11	100-00103	BL/V2	NOTDETEC			
12	100-00103	V3	NOTDETEC			

3. Running the Application

3.1 Understanding the user interface

*Users must first have access to the Shiny Server – requires Lilly username and password.
**This application lacks some functionality in Internet Explorer – Google Chrome is highly recommended.

Once the link has loaded, this is the landing page (Table tab) that will appear.
Elements of the primary interface include the following:
1) **File input field.** When clicked, the Browse button will open File Explorer.
2) **Checkboxes.** When selected, these will show or hide text input fields for subsequent columns containing tier results.

3) **Text input field**. The user is required to complete all entry boxes with the corresponding, case-sensitive values for a loaded dataset.
4) **Numeric input field.** Sets the MRD for this particular dataset and is used for building many tables and summary statistics. Can be altered by the user to reactively update the tables and statistics. Default is set to 10.
5) **Selection input field for visit codes.** Once the baseline visit code has been filled out, this field loads all unique visit codes from the dataset, plus columns for Subject and highest titer per subject. It allows the user to rearrange the pivot tables in the best chronological order that they see fit. This also sets the X-axis for the time series graph (Section 3.3).
6) **Dropdown list.** Shows more tables, many of which are reactively built once the "Baseline Visit" value has been entered. Others will require more user input.
7) **Download button.** Creates an export file that includes the original dataset along with all additional tables created in the app.

3.2 First tab: Table

As a file is loaded into the app, the main panel will reactively display the table.

Load a dataset:

Browse... AMAC_PBI_trimme
Upload complete

Select all columns that appear in the dataset:
- ■ Tier 1
- ■ Tier 2   ☐ Tier 2b
- ■ Tier 4   ☐ Tier 4b

Complete the fields below:

BL/V2

DNR

NOTDETEC

DETECTED

NOTDETEC

DETECTED

NOTDETEC

Enter Minimum Required Dilution:

10

MRD set at 1:10

| | Table | Flags | Plot | Summary | Help | | |

Show 10 ▼ entries   Search: [   ]

Subject	Visit	Tier1	Tier2	Tier3	Tier4
100-00102	Baseline	NOTDETEC			
100-00102	V3	NOTDETEC			
100-00102	V4	NOTDETEC			
100-00102	V6	NOTDETEC			
100-00102	V7	NOTDETEC			
100-00102	EV1	NOTDETEC			
100-00102	EV2	DNR	DETECTED	10	NOTDETEC
100-00102	EV3	DNR	DETECTED	10	NOTDETEC
100-00102	EV6	NOTDETEC			
100-00103	Baseline	NOTDETEC			

Showing 1 to 10 of 3,404 entries

Previous **1** 2 3 4 5 ... 341 Next

When the table is loaded on the page, the user should begin completing items on the side panel as it pertains to their data. These include values for baseline visit, and detected and not detected values for all tiers of testing. In the example above, the dataset contains columns for Tiers 1-4. Thus, only the corresponding checkboxes for Tiers 1, 2, and 4 need to be selected. These checkboxes will reveal additional text fields for detected and not detected values in each tier. Tier 3 values should be numeric, and most calculations around this column will function from the baseline text entry and the numeric input for MRD.

Text fields are case-sensitive – capitailization, spacing, and spelling must match the values in the table. Copy and pasting values from the table to the text fields works fine, but double check for uneccessary white spaces in the text input field.

The "Select a view" dropdown shows seven tables that are available for viewing. Most will become accessible after the text input fields are completed:

1) **Original:** The initial table that was loaded into the tool. The titer column is processed automatically, trimming off every instance of the string "1:" if identified.
2) **Baselines:** A subset of the **Original** table, only displaying rows that match the user-entered baseline visit value for this dataset.
3) **Baseline Positives:** A subset of the **Original** table, only displaying rows that a) match the user-entered baseline visit value and b) match the user-entered Tier2 detected value for this dataset.
4) **Unevaluated Subjects:** Displays subjects who either a) are missing the baseline visit value and/or b) have a baseline visit without any follow-up visits. If the dataset does not logically identify any rows for either of these checks, the table is populated with an "EMPTY" premise. Subjects that appear in this table are not considered in the calculations for treatment emergence (Section 3.5).
5) **Subject Pivot Table:** Displays each unique subject and their Tier3 values for each visit. The highest post-baseline titer for each subject is also appended to this table (maxTiter). Any value of "0" indicates that the visit occurred for that Subject but no titer was reported. Any value of "–" is a placeholder, indicating that a visit for that subject has not occurred.
6) **Treatment Emergent Pivot Table:** A subset of the **Subject Pivot Table**, only displaying the following rows:
   a. Treatment-induced subjects: tested negative at baseline and have a post-baseline reported titer that is at least two times greater than the MRD
   b. Treatment-boosted subjects: tested positive at baseline and have a post-baseline reported titer that is at least four times greater than the baseline titer
7) **Titer Pivot Table:** Shows each unique baseline titer and the counts of each maximum post-baseline titer. Any value of "0" indicates a titer was not detected. Grand totals of each row and column are included.

*The input field "Reorganize Visit Codes" only applies to the **Subject Pivot Table**, **Treatment Emergent Pivot Table**, and the X-axis for the time-series plot.*

It is possible to have data that yields no results for baseline positives and/or treatment emergent subjects. These tables are handled in a similar way to the **Unevaluated Subjects** table, populating the tables with feedback that indicates no subjects were identified.

The Download All Tables button provides a quick and simple export to an Excel workbook. All seven tables in the tool are separated into distinct Excel sheets. An eighth sheet is also included for any search results that a user completes in the search bar. In order for search results to appear in the Excel workbook, the search must be done on the **Original** table view.

3.3 Second tab: Flags

This QC table updates with each text input entry provided by the user. Logical statements are built into the tool to identify various types of QC checks from Tier 1 through Tier 4.

The "Premise" column provides a description as to why each row is flagged. A frequency table showing the sum of each Premise in the QC table is also included.

The QC checks can be categorized in the following ways:

- **Discrepant value flags:** instances where any of the values in the tier columns do not match the user-entered values for detected and not detected. In Tier3, values that are not a multiple of the MRD or are non-numeric are also flagged.

- **Between tier flags:** logical scenarios where test results are inconsistent or unnecessary from one tier to the next.
- **Duplicate visit flags:** shows subjects who have multiple records of the same visit code (it is possible that these are retests).

An exhaustive list of all possible QC checks can be found on the Help tab.

3.4 Third tab: Plot

*The titer histogram reactively charts the frequency of the highest titer in the **Subject Pivot Table**. It automatically excludes a bin for subjects who have no reported titers (maxTiter = 0 in pivot table).*

The total number of highest post-baseline titers is displayed beneath the table. It is worth noting that instances where subjects missed baseline but later had a reported titer in the **Subject Pivot Table** are also reported in this plot. However, subjects who miss baseline are not considered for treatment emergent. Additionally, it is possible that this histogram also includes high post-baseline titers from subjects who do not become treatment emergent.
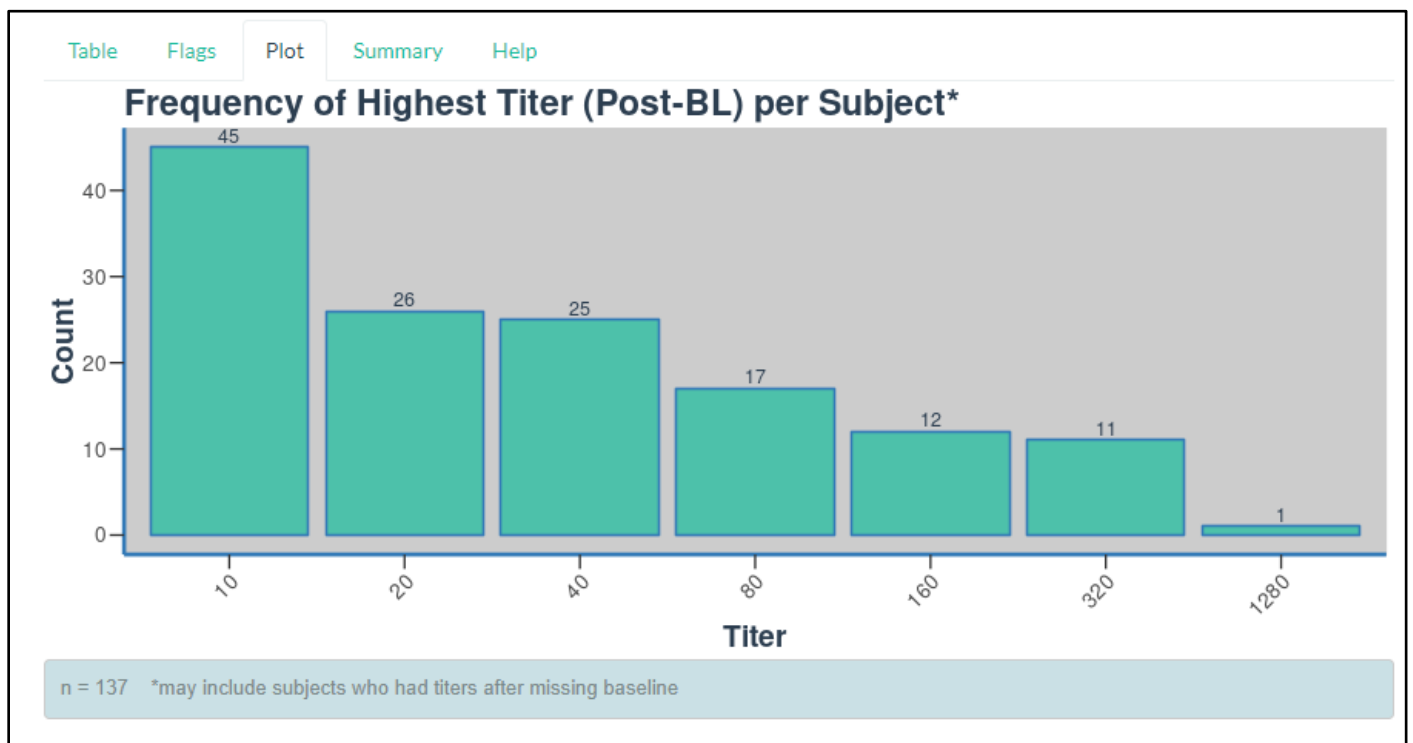
*The time series line graph reactively plots two numeric variables from one unique subject in the dataset. This could be any combination of data from ADA, PK/PD, Biomarker, etc. It is imperative to include the **Subject** and **Visit** selections in the second field to maintain proper functionality.*

Users are free to choose two numeric variables to plot against one another. Besides Tier3, column names are insignificant. The numeric input field titled **Multiplier for Y-Axis Scale** allows for one y-axis to be transformed as a multiple of the other. This field defaults to 1 but can be set to any number, including decimals less than 1.
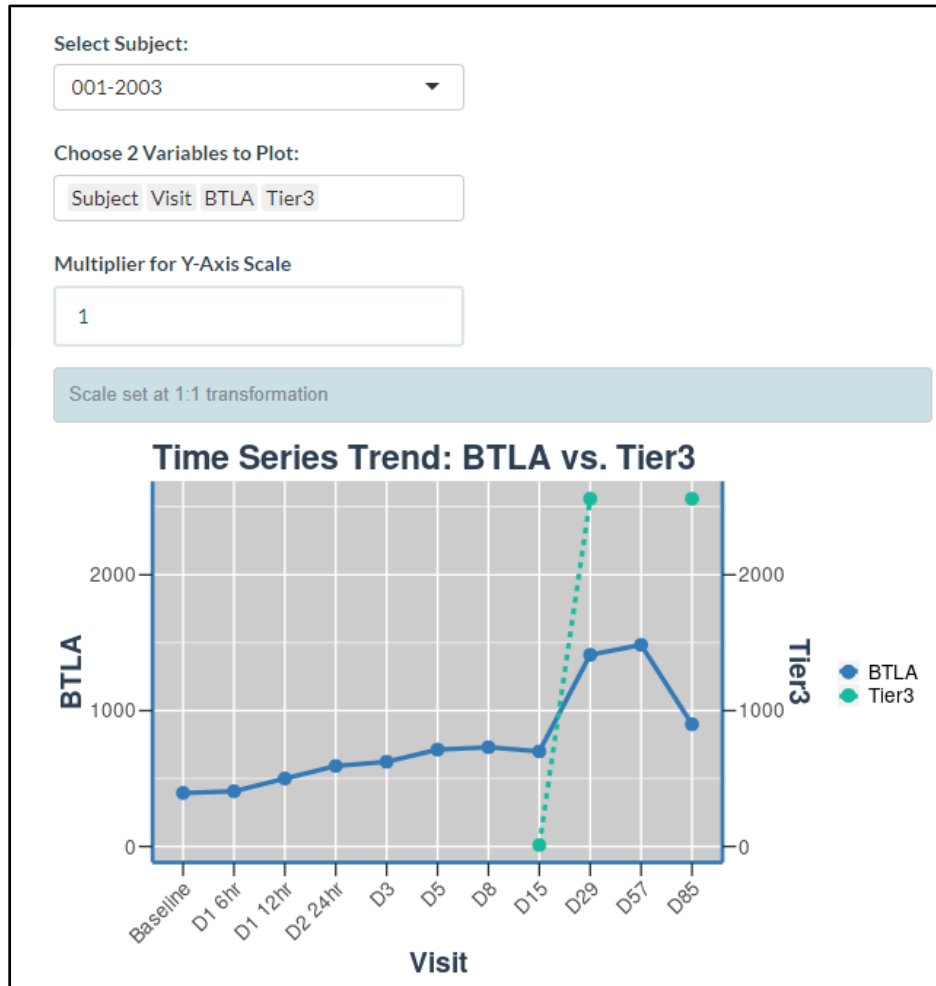
This graph will not plot Subjects without at least one instance of a numeric value for either of the 2 selected variables. For example, if the user tries to plot a subject without any reported titers AND without any reported Tier 4 values, the graph will throw an error because there is no numeric data to plot.

Remember, the selection input field for visit codes on the side panel will help rearrange any Visits along the X-axis that are not plotted in chronological order.

3.5 Fourth tab: Summary

A few tables reactively update and display statistics once the user begins filling out the text input fields. The second and third tables will be built upon the user entering a value for baseline visit.

The first table displays statistics for samples in each tier. This table automatically identifies unique tier columns in the dataset and creates a row for each tier.
1) **SamplesTested** is the sum of the user-entered detected and not detected values in each tier.
2) **Detected** is the sum of the user-entered detected values in each tier.
3) **PositiveRate** is # samples **Detected** / # **SamplesTested** in each tier. Putative positive rate and confirmed positive rate correspond to Tier1 and Tier2 respectively.

Table	Flags	Plot	Summary	Help

	TierSubset	SamplesTested	Detected	PositiveRate
Total				
	Tier 1	3403	1324	38.91%
	Tier 2	1324	886	66.92%
	Tier 4	886	532	60.05%
Baseline				
	Tier 1	249	66	26.51%
	Tier 2	66	38	57.58%
	Tier 4	38	15	39.47%
Post-Baseline				
	Tier 1	3154	1258	39.89%
	Tier 2	1258	848	67.41%
	Tier 4	848	517	60.97%

	Count	Rate
Evaluable Subjects for Treatment Emergence	247	97.63%
Evaluable Subjects w/ Baseline Positive Result	38	15.38%
Treatment Emergent Subjects	75	30.36%
Treatment-Induced	59	23.89%
Treatment-Boosted	16	6.48%

	Sort ⬍
Total Unique Subjects	253
Unevaluated Subjects	6
Highest Baseline Titer	320
Highest Post-Baseline Titer	1280

The second table displays treatment emergent statistics.

1) **Evaluable Subjects for Treatment Emergence** are those that have a baseline visit and at least one follow-up visit.
    a. **# Evaluable Subjects / # Total Subjects = % of Evaluable Subjects Rate**
2) **Evaluable Subjects w/Baseline Positive Result** are detected in Tier2 at the Baseline visit.
    a. **# Baseline Positive Results / # Evaluable Subjects = % of Baseline Positive Rate**
3) **Treatment Emergent Subjects** are those that are treatment emergent and are either:
    a. **Treatment-Induced**: subject has an unreported titer at baseline but reaches a titer that is two times the MRD at a later visit.
    b. **Treatment-Boosted**: subject has a reported titer at baseline and reaches a titer that is four times their baseline titer at a later visit.

The third table displays a few more general statistics about the dataset.

1) **Total Unique Subjects** is the sum of **Evaluable Subjects** and **Unevaluated Subjects**. It is the number of distinct patients in the dataset.
2) **Unevaluated Subjects** are those that either:
    a. Missed baseline visit
    b. Have a baseline visit without any follow-up visits
3) **Highest Baseline Titer** is the maximum Tier3 value from Baseline visits only
4) **Highest-Post Baseline Titer** is the maximum Tier3 value from all visits excluding Baseline

3.6 Fifth tab: Help

Users can consult this page to get a synopsis on a few main ideas, such as how to preprocess a file correctly, get a detailed list of all logical QC checks built into the tool, and a recap of how the summary statistics tables are created.

4. Application Limitations

4.1 Known errors

Initially, the web page will display a few errors on each tab if data has not been loaded into the tool. However, these errors could continue to appear for a number of reasons.

1) Incorrectly formatted column headers. Refer to Section 2.2 to troubleshoot a dataset.
2) Missing required columns. In order to obtain the bare minimum functionality, a few columns must be present and correctly formatted in every dataset. Refer to Section 2.2 to review specific syntax.

3) Datasets that yield 0 treatment emergent subjects will show an empty table on the **Treatment Emergent Pivot Table**. The second table on the Summary tab, displaying treatment emergent statistics, could have errors or be missing entirely.
4) By default, the R Shiny application will disconnect from the server after 15 minutes of inactivity.
5) By default, the R Shiny application has a maximum file upload limit of 5MB.
6) By default, the R Shiny application has a maximum timeout at 45 seconds for downloading files.