

SESUG Paper 47-2024

Streamlining Vaccine Study Output Generation with a Versatile Efficacy Macro

Zhengyang Shi, Ying Zhang Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

With the increasing prominence of vaccine studies with efficacy analysis, the need for standardized output generation and improved programming efficiency becomes critical. A vaccine efficacy macro has been developed supporting multiple methods and dynamic control over output formats and information. The paper provides an overview of the vaccine efficacy macro's design, output structure, highlight key features and provides an example macro call.

INTRODUCTION

With the growing number of vaccine studies with assessment of effectiveness, streamline the process of analyzing the efficacy of vaccine and improve programming efficiency become imperative. To address this need, a vaccine efficacy macro has been developed. The efficacy macro provides multiple statistical methods, scenarios, and dynamic control of the outputs. Additionally, it generates output in a standard format and offer an efficient solution that can save time and effort. The paper aims to provide a comprehensive understanding of the macro's design, structures, and parameters.

THE MACRO DESIGN

To utilize the efficacy macro, there are some require specific variables in the input Efficacy Analysis Dataset (ADEFF). These flags serve two purposes, one is to identify the records used in the efficacy macro to display different information, second is to enable traceability that reviewer easily track which records are using to display or calculate in the outputs.

The analysis flags will indicate one record per subject per combination scenarios display in the outputs, for example,

- For endpoint row, each endpoint or PARAMCD will have one record per subject being flagged in the dataset.
- For subgroup level rows, under each endpoint and each level of subgroup will have one record per subject being flagged in the dataset. For example, ADEFF has two endpoints, and each endpoint has one subgroup with two levels. So, every subject will have four analysis flags (2*2).

In addition to analysis flags, the dataset should also include following variables,

- Analysis Day (ADY): derived as date of becoming a case or the latest follow-up date - user-specified follow-up period start date +1 and will be directly used in the efficacy macro as the subject follow-up time.
- Case indicator (AVAL): case status; 1=case and 0=non-case.

THE KEY PARAMETERS

Subject-Level Analysis Dataset (ADSL) and Efficacy Analysis Dataset (ADEFF) are the two-input datasets. User could select population for the analysis.

For the endpoint levels. User could input multiple endpoints, its label and analysis flag from efficacy analysis dataset. Corresponding parameters are ENDPT, analysis endpoint(s) from efficacy analysis dataset, if there are multiple analysis endpoints, separate them by (/). ENDPTLBL, analysis endpoint labels, Label(s) for analysis endpoint(s) displayed in the output. Required when user selects to display the overall analysis row for the endpoint, i.e., display_overall_row=Y. If there are multiple analysis endpoints, separate them by (/). ENDPT_ANALYSISFLAG, Endpoint analysis flag, required when user selects to display the overall analysis row(s) for analysis endpoint(s), i.e., display_overall_row=Y. Only one analysis flag can be specified and will be used for all endpoints.

The efficacy macro supports to display up to two subgroups under each endpoint. Similar to endpoint, user is required to specify the subgroup variables, labels, and analysis flags. Additionally, it allows to display the combined subgroup analysis whenever there is an interest in subcategory of serotype, which user will specify label and analysis flag for each combined subgroup analysis. For example, serotype A+B+C or serotype A+D.

Besides input main part of the outputs, the efficacy macro also let user control the treatment group, title, footnote, decimal point for different column, font size, confidence interval, minimum n to display confidence interval display and display negative vaccine efficacy or not.

The last part is for statistical methods. The efficacy macro offers users the flexibility to select from three methods for calculating vaccine efficacy (VE). Once parameter METHOD is chosen, user could provide additional details specific to each method.

If method selected is exact binomial method¹, user can decide the type of confidence interval using parameter CI_TYPE and have the option to choose clopper Pearson (CP) or Blaker confidence interval.

If method selected is modified Poisson regression with robust variance⁴, user can specify categorical and/or continuous covariates in the modified Poisson regression. Separate covariates by (/). Categorical or continuous covariates will be determined by variable format: character format will be defined categorical and numerical format will be defined as continuous.

If method selected is group sequential method³, user has to specify the STAGE (The nth analysis for the group sequential method), TOTCASE (total observed cases for each stage), futility_boundary (Futility boundary for each stage, the minimum cumulative case number in the treatment group to claim futility at each stage. Separate the boundary for each stage by (/)) and efficacy_boundary, (Efficacy boundary for each stage, The maximum cumulative case number in the treatment group to claim efficacy at each stage. Separate the boundary for each stage by (/)). And the P-value for the efficacy analysis is calculated using the exact method using binomial CDF².

OUTPUT

The efficacy macro has the capability to generate a comprehensive output as shown below. The output has following information.

- The header includes the number of randomized and vaccinated subjects for each treatment and column of total number of cases.
- Under each treatment, it has total number of cases, number of subjects contributed to the efficacy analysis, user defined follow up and incidence rate estimate per user defined period associate their confidence interval (CI).
- The last section contains the columns with the estimate, confidence interval (CI) and p-value between treatment based on user select methods.

User could customize any column to be displayed in the output and the wording in the header for each column. For example, user could only display the incidence rate column without the CI, only display the within treatment information or the observed efficacy p-value only in endpoint row.

Analysis of Vaccine Efficacy
(Subtitle 1)
(Subtitle 2 - Analysis Population)

Endpoint	Total Number of Cases	Treatment 1 (N=xxx)				Treatment 2 (N=xxx)				Observed Efficacy (%)	
		n	Number of Cases	Person-years Follow-up ^a	Incidence Rate Estimate per 100 Person-years ^b (95% CI) ^c	n	Number of Cases	Person-years Follow-up ^a	Incidence Rate Estimate per 100 Person-years ^b (95% CI) ^c	Estimate (95% CI) ^d	p-Value ^d (1-sided)
Endpoint 1	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
By Subgroup1 Type											
Subgroup1 1	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup1 2	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup1 3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup1 1&3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup1 2&3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
By Subgroup2 Type											
Subgroup2 1	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup2 2	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup2 3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup2 1&3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
Subgroup2 2&3	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx

^a Year is calculated as 365 days.
^b 100 person-years is calculated as 36500 days.
^c Estimated based on the exact method.
^d Estimate and 95% CI are based on the exact binomial method, and 1-sided p value is based on the exact method.
N=Number of participants randomized and vaccinated, n=Number of participants contributing to the analysis.
CI=Confidence interval.

Same as the column adjustment, User has the flexibility to adjust the row information in the efficacy macro. It is designed to handle multiple endpoints, allowing users to display up to two subgroups under each endpoint.

From the simplest scenario of displaying multiple endpoints to most complex scenarios such as displaying multiple endpoints with two subgroup each plus combination of subgroup. The efficacy macro offers the flexibility for users to select any row combination based on their interests and study requirements.

Additionally, the efficacy macro supports analysis without showing endpoint level and provides flexibility to display only the subgroups levels or timepoint on their specific needs. it allows user to choose cumulative or incremental analysis based on how the user define ADY.

User also has the flexibility to choose different statistical methods for endpoint row compared to subgroup rows. For instance, they could choose modified Poisson for endpoint row, while using exact binomial for subgroup rows. Alternatively, user could apply same method but with different setting, for instance, using exact binomial with CP confidence interval for endpoint row and selecting exact binomial with blaker confidence interval for subgroup rows. Similarly, user could apply the modified Poisson for both the endpoint and subgroup rows but with different covariates.

MACRO FEATURES

KEY FEATURES OF THE MACRO INCLUDE THE FOLLOWING

- Multiple statistical methods for vaccine efficacy – user has the option to choose and define certain parameters in three methods. Moreover, user could specify different methods between endpoint row and subgroup rows.
- Satisfy most analysis needs – the efficacy macro could support overall analysis, subgroup analysis and analysis by timepoint/serotype.

- Flexible to customize the outputs per user needs – the user has the option to choose how and what to display in the horizontal and vertical way in the output.
- The macro could handle multiple endpoints and up to two subgroups under each endpoint, including by timepoint with specifying the start of time intervals and analysis by subgroup only.

EXAMPLE OF A MACRO CALL

The example call has three sections.

- The left section has the input analysis dataset, one endpoint with one subgroup. Users specify the endpoint parameter, its label and analysis flag, and related subgroup variable, its label and analysis flag.
- The middle section contains user input on treatment information, follow up time unit, incidence rate unit and Ve_0 -- Vaccine efficacy assumption under the null hypothesis. ve_neg is Y means display negative VE estimate when it is negative. Additionally, Users select exact method with black CI.
- The right section that user select to display endpoint row, total case column and incidence rate CI column. When user select the display efficacy column with overall, the p-value will only display in the endpoint level. Additionally, users input decimal point and title for the output.

```
%vr0ve0anal(
population_from = lptda.adsl,
observation_from = lptda.adeff,
population_where = %str(RANDFL='Y' and FASFL='Y'),
endpt = MALRI150,
endptlbl = RSV HOSP Days 1-150,
endpt_analysisflag= ANL01FL,
subgroup1_variable= rsvtype,
subgroup1_level=all,
subgroup1_label=Sub,
subgroup1_analysisflag=ANL02FL,
subgroup1_combine_analysisflag=,
subgroup1_combine_label=,
subgroup2_variable=,
subgroup2_level=,
subgroup2_label=,
subgroup2_analysisflag=,
subgroup2_combine_analysisflag=,
subgroup2_combine_label=,

header_label=,
trtgrp = TRT01P,
treatment = %str(Treatment A),
control = %str(Treatment B),
fuptime_unit = year,
ir_unit = %str(100 Person-years),
fuptime_unit_df= 365,
ci_level =.95 ,
ve_0 = 0.25,
n_cutoff = ,
ve_neg = Y,
method =exact ,
ci_type =blaker,
covariate=,
stage= ,
totcase= ,
efficacy_boundary = ,
futility_boundary = ,

display_overall_row=Y,
display_total_case=Y,
display_ir_ci=Y,
display_efficiency_column=overall,
decimal_fuptime=1,
decimal_ir=1,
decimal_pvalue=3,
title_text_1= Analysis of Vaccine Efficacy,
title_text_2= (Subtitle 1),
title_text_3= (Subtitle 2 - Analysis Population),
end_notes=,
font_size=9,
page_size=,
data_source_txt=%str([ adam-adsl; adeff]) ,
output_name=example1,
create_output_dataset=Y,
pagesize_adjust=%str(+10),
```

Below is the example output with user choice on title, footnote, and column display. The efficacy macro could adopt different scenarios that fit the user needs.

Analysis of Vaccine Efficacy (Subtitle 1) (Subtitle 2 - Analysis Population)											
Endpoint	Total Number of Cases	Treatment A (N=xxx)				Treatment B (N=xxx)				Observed Efficacy (%)	
		n	Number of Cases	Person-years Follow-up ^a	Incidence Rate Estimate per 100 Person-years ^b (95% CI) ^c	n	Number of Cases	Person-years Follow-up ^a	Incidence Rate Estimate per 100 Person-years ^b (95% CI) ^c	Estimate (95% CI) ^d	p-Value ^d (1-sided)
RSV HOSP Days 1-150	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	x.xxx
By Sub Type											
rsvtype1	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	
rsvtype2	xxx	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xxx	xxx	xxx.x	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	

CONCLUSION

The vaccine efficacy macro is a valuable and versatile tool that allows user to customize their outputs according to their requirement. One of the advantages of this macro is to provide users with control both the information and the formatting of the outputs. This capability allows programmer to generate efficacy results in a standard and consistent way, ensuring accuracy and efficiency.

REFERENCES

1. Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28(4), 783-798.
2. Chan, I. S., & Bohidar, N. R. (1998). Exact power and sample size for vaccine efficacy studies. *Communications in Statistics-Theory and Methods*, 27(6), 1305-1322.
3. Jennison, C., & Turnbull, B. W. (1999). *Group sequential methods with applications to clinical trials*. CRC Press.
4. Zou, G. (2004). A modified poisson regression approach to prospective studies with binary data. *American journal of epidemiology*, 159(7), 702-706.