PharmaSUG 2020 - Paper SA-051

Calculation of Cochran–Mantel–Haenszel Statistics for Objective Response and Clinical Benefit Rates and the Effects of Stratification Factors

Girish Kankipati and Chia-Ling Ally Wu, Seattle Genetics, Inc.

ABSTRACT

In oncology clinical trials, primary and secondary endpoints are analyzed using different statistical models based on the study design. Objective response rate (ORR) and clinical benefit rate (CBR) are commonly used as key endpoints in oncology studies, in addition to overall survival (OS) and progression-free survival (PFS). The use of ORR and CBR as an endpoint in these trials is widespread as objective response to therapy is usually an early indication of treatment activity and it can be assessed in smaller samples compared to OS; furthermore, FDA considers ORR and CBR as clinical and surrogate endpoints in traditional and accelerated approvals.

Bringing new therapies to market based on ORR and CBR requires specialized statistical methodology that not only accurately analyzes these key endpoints but can also accommodate stratified study designs aimed at controlling for confounding factors. The Cochran-Mantel-Haenszel (CMH) test provides a solution to address these many needs.

This paper introduces CMH test concepts, describes how to interpret its statistics, and shares insights into SAS® procedure settings to use it correctly. The calculation of ORR and CBR with 95% confidence intervals using the Clopper-Pearson method and strata-adjusted p-values using the CMH test are discussed with sample data and example table shells, along with examples of how to use the FREQ procedure to calculate these values.

INTRODUCTION

Objective response rate (ORR) and clinical benefit rate (CBR) are commonly used as key endpoints in oncology clinical trials. FDA guidance for cancer clinical trial endpoint evaluations has mentioned tumor assessments, including ORR, can support overall survival (OS) as surrogate endpoints for traditional approval and accelerated approval (U.S. Food and Drug Administration, 2018). The advantages of deriving ORR and CBR are that they can be assessed earlier than PFS and OS on during a clinical trial, require a smaller sample size compared to survival analysis, and the effect is attributable to treatment instead of the natural progression of disease. ORR and CBR can be used to help predict the success of treatment early in some situation.

In most cases, we may deal with ORR and CBR point estimates and 95% confidence intervals and not as much with how they are evaluated for statistical significance, especially in stratified analyses to control confounders in the study. Hence, this paper aims to give an understanding of the accurate method of using ORR and CBR in a stratified analysis and observing the effect of the stratification factors.

OBJECTIVE RESPONSE RATE (ORR)

The FDA definition of ORR is "the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period" (U.S. Food and Drug Administration, 2018). Generally, the FDA has defined ORR as the sum of complete responses (CRs) and partial responses (PRs). This implies ORR is a direct measure of a drug antitumor activity, which can be evaluated in a single-arm study. The significance of ORR is assessed by its magnitude and duration and the percentage of CRs. It can also be a surrogate endpoint to support both accelerated and traditional marketing approval. Therefore, ORR is a common endpoint in oncology clinical trials.

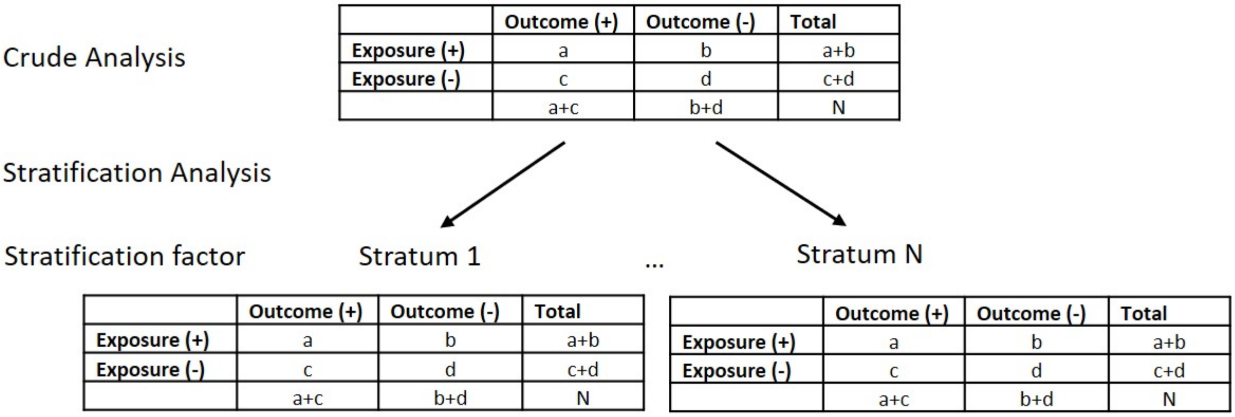
CLINICAL BENEFIT RATE (CBR)

CBR is defined as the percentage of patients with advanced or metastatic cancer who have achieved complete response, partial response, and stable disease while on a therapeutic intervention in clinical trials of anticancer agents. The frequent use of these measures of drug efficacy presents the question of whether CBR is a useful additional endpoint in early clinical trials, and if it can reasonably predict the success of an agent in subsequent, adequately powered, randomized trials.

COCHRAN-MANTEL-HAENSZEL TEST

The Cochran-Mantel-Haenszel (CMH) test (or Mantel-Haenszel) is an inferential test for the association between two binary variables, while controlling for a third confounding nominal variable (Cochran 1954; Mantel and Haenszel 1959; Paul 2017). It is used to generate an estimate of an association between an exposure and an outcome after taking into account confounding factors. Essentially, the CMH test examines the weighted association of a set of exposure-by-outcome tables. The most common situation is multiple tables of independence (for example, exposure-by-outcome table as below Figure 1) and performing the experiment multiple times.

Figure 1 demonstrates how to estimate the association by crude analysis and stratification analysis. Crude analysis uses a standard layout for the exposure by outcome table to show unadjusted associations, while stratification analysis separates the crude analysis to a series of strata driven by the number of confounder categories, calculates the statistics in each stratum, and uses CMH test to derive the stratification-adjusted association. For example, if we want to analyze the association between treatment and patient survival status after one year of treatment while considering gender as a confounder, the crude analysis only considers treatment and survival status, whereas the stratification analysis has treatment and survival status in male and female strata.



Crude Analysis

	Outcome (+)	Outcome (-)	Total
Exposure (+)	a	b	a+b
Exposure (-)	c	d	c+d
	a+c	b+d	N

Stratification Analysis

Stratification factor

Stratum 1

	Outcome (+)	Outcome (-)	Total
Exposure (+)	a	b	a+b
Exposure (-)	c	d	c+d
	a+c	b+d	N

...

Stratum N

	Outcome (+)	Outcome (-)	Total
Exposure (+)	a	b	a+b
Exposure (-)	c	d	c+d
	a+c	b+d	N

Figure 1. Crude Analysis vs. Stratification Analysis

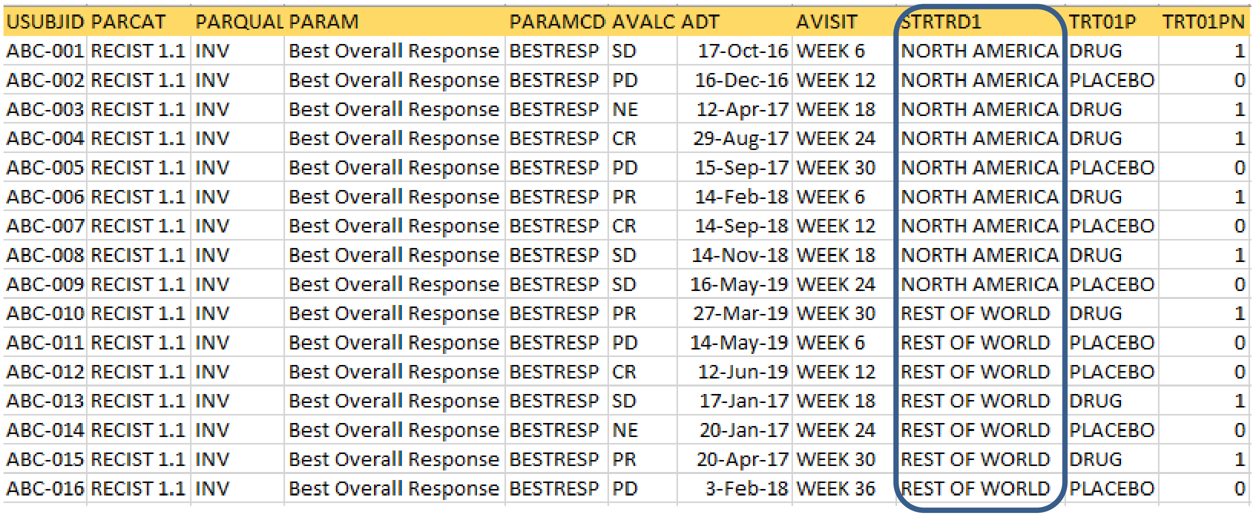
In crude analysis, a chi-square or Fisher's exact test could be appropriate tests depending on the expected value in each cell. Using a stratification analysis to control confounding, a CMH test is appropriate to determine the adjusted association between exposure and outcome. Below is the formula to calculate the CMH statistic.

$$x_{MH}^2 = \frac{\{\,|\sum\,[\,a\;-\;(a+b)\,(a+c)/n]\;|\;-\;0.5\;\}^2}{\sum(a+b)(a+c)(b+d)(c+d)/(n^3-n^2)}$$

The numerator is the squared sum of deviations between the observed and expected values under the null hypothesis. The observed value is the value in cell a, while the expected value is (a+b)(a+c)/n, and the 0.5 is a continuity correction. The denominator contains an estimate of the variance of the squared differences. The test statistic, x_{MH}^2, is chi-square distributed with one degree of freedom used to calculate the p-value.

EXAMPLE DATASET

To illustrate this topic, we constructed an example dataset as a dummy CDISC ADaM ADRS (response analysis) data set, as shown in Table 1. The parameter “Best Overall Response” is a subject’s best response value stored in variable AVALC, which is evaluated based on RECIST 1.1. AVALC contains: CR (Complete Response), PR (Partial Response), SD (Stable Disease), PD (Progressive Disease), and NE (Not Evaluable). Variable STRTRD1 consists of region: North America, and Rest of World. Treatment information is stored in TRT01P and TRT01PN, Drug (1) vs. Placebo (0).



USUBJID	PARCAT	PARQUAL	PARAM	PARAMCD	AVALC	ADT	AVISIT	STRTRD1	TRT01P	TRT01PN
ABC-001	RECIST 1.1	INV	Best Overall Response	BESTRESP	SD	17-Oct-16	WEEK 6	NORTH AMERICA	DRUG	1
ABC-002	RECIST 1.1	INV	Best Overall Response	BESTRESP	PD	16-Dec-16	WEEK 12	NORTH AMERICA	PLACEBO	0
ABC-003	RECIST 1.1	INV	Best Overall Response	BESTRESP	NE	12-Apr-17	WEEK 18	NORTH AMERICA	DRUG	1
ABC-004	RECIST 1.1	INV	Best Overall Response	BESTRESP	CR	29-Aug-17	WEEK 24	NORTH AMERICA	DRUG	1
ABC-005	RECIST 1.1	INV	Best Overall Response	BESTRESP	PD	15-Sep-17	WEEK 30	NORTH AMERICA	PLACEBO	0
ABC-006	RECIST 1.1	INV	Best Overall Response	BESTRESP	PR	14-Feb-18	WEEK 6	NORTH AMERICA	DRUG	1
ABC-007	RECIST 1.1	INV	Best Overall Response	BESTRESP	CR	14-Sep-18	WEEK 12	NORTH AMERICA	PLACEBO	0
ABC-008	RECIST 1.1	INV	Best Overall Response	BESTRESP	SD	14-Nov-18	WEEK 18	NORTH AMERICA	DRUG	1
ABC-009	RECIST 1.1	INV	Best Overall Response	BESTRESP	SD	16-May-19	WEEK 24	NORTH AMERICA	PLACEBO	0
ABC-010	RECIST 1.1	INV	Best Overall Response	BESTRESP	PR	27-Mar-19	WEEK 30	REST OF WORLD	DRUG	1
ABC-011	RECIST 1.1	INV	Best Overall Response	BESTRESP	PD	14-May-19	WEEK 6	REST OF WORLD	PLACEBO	0
ABC-012	RECIST 1.1	INV	Best Overall Response	BESTRESP	CR	12-Jun-19	WEEK 12	REST OF WORLD	PLACEBO	0
ABC-013	RECIST 1.1	INV	Best Overall Response	BESTRESP	SD	17-Jan-17	WEEK 18	REST OF WORLD	DRUG	1
ABC-014	RECIST 1.1	INV	Best Overall Response	BESTRESP	NE	20-Jan-17	WEEK 24	REST OF WORLD	PLACEBO	0
ABC-015	RECIST 1.1	INV	Best Overall Response	BESTRESP	PR	20-Apr-17	WEEK 30	REST OF WORLD	DRUG	1
ABC-016	RECIST 1.1	INV	Best Overall Response	BESTRESP	PD	3-Feb-18	WEEK 36	REST OF WORLD	PLACEBO	0

Table 1. ADRS Dataset with Stratification Variable (North America vs. Rest of World)

EFFECT OF STRATIFICATION FACTORS ON ORR AND CBR

Prior to digging into a stratification analysis, we wanted to observe if region confounds the association between treatment and ORR, e.g., whether patients in different regions who receive the same treatment may have a different outcome in ORR. The approach presented in Figure 1 compares the crude association with the associations in each stratum stratified by confounding factor. From the data we have the unadjusted odds ratio (OR), the OR in North America, and the OR in Rest of World. The results are presented in Table 2, showing the ORs are apparently different between these two regions both in point estimate and in the 95% confidence interval (CI), which means the likelihood of achieving ORR differs by region. Therefore, the region adjustment for treatment and response is necessary.

Odds Ratio and Relative Risks

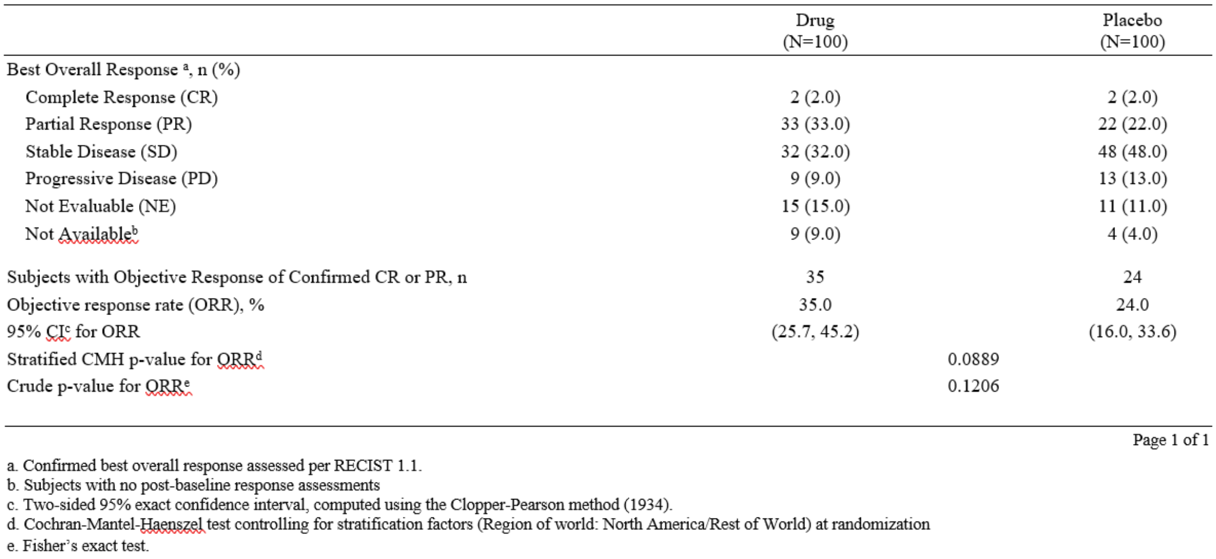
Statistic	Value	95% Confidence Limits	
Odds Ratio (Unadjusted)	0.4068	0.2314	0.7214
Odds Ratio in North America	0.6169	0.2799	1.3597
Odds Ratio in Rest of World	0.2654	0.1160	0.6071

Table 2. Odds Ratio of Achieving ORR in Crude and Stratified by Region

In the examples below, we are going to discuss the effect of stratification by region on the association between exposure and response, using ORR and CBR as the example endpoint. We will present a model analysis with and without stratification and briefly discuss the SAS® code to conduct the analysis.

OBJECTIVE RESPONSE RATE

Objective response is defined as the best overall response of complete (CR) or partial response (PR) using RECIST 1.1. ORRs were calculated on our example data and the results are shown below in Figure 2.



	Drug (N=100)		Placebo (N=100)
Best Overall Response [a], n (%)			
Complete Response (CR)	2 (2.0)		2 (2.0)
Partial Response (PR)	33 (33.0)		22 (22.0)
Stable Disease (SD)	32 (32.0)		48 (48.0)
Progressive Disease (PD)	9 (9.0)		13 (13.0)
Not Evaluable (NE)	15 (15.0)		11 (11.0)
Not Available[b]	9 (9.0)		4 (4.0)
Subjects with Objective Response of Confirmed CR or PR, n	35		24
Objective response rate (ORR), %	35.0		24.0
95% CI[c] for ORR	(25.7, 45.2)		(16.0, 33.6)
Stratified CMH p-value for ORR[d]		0.0889	
Crude p-value for ORR[e]		0.1206	

Page 1 of 1

a. Confirmed best overall response assessed per RECIST 1.1.
b. Subjects with no post-baseline response assessments
c. Two-sided 95% exact confidence interval, computed using the Clopper-Pearson method (1934).
d. Cochran-Mantel-Haenszel test controlling for stratification factors (Region of world: North America/Rest of World) at randomization
e. Fisher's exact test.

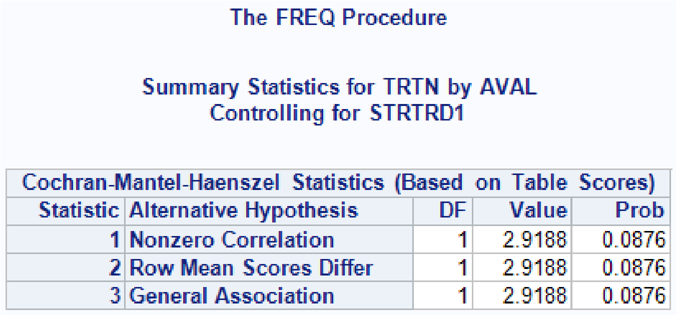
Figure 2. Example Output for the Objective Response Rate

SAS® PROC FREQ can perform the stratification analysis as per the code below:

```
ods listing close;
ods output cmh = <output dataset>;
proc freq data = <source dataset>;
   tables strtrd1*trt01pn*aval/cmh;
run;
ods listing;
```

We know the SAS® FREQ procedure is used to produce n-way by n-way crosstabulation tables. In our stratification analysis, we requested an association between treatment (variable TRT01PN) and outcome (variable AVAL), adjusted by region (variable STRTRD1), using CMH as the statistical method.The value of variable AVAL is 1 for the subjects with ORR and 0 without ORR.

The option CMH in the TABLES statement is used to request CMH statistics. The ODS OUTPUT statement produces a SAS dataset from the portion of the listing produced by PROC FREQ that corresponds to output produced by the option used in the TABLES statement, in our case CMH. From the OUTPUT list, CMH produces correlation, association statistics, CMH adjusted odds ratios and relative risks, and Breslow-Day test results. In our analysis, we took the general association under the CMH statistics.



The FREQ Procedure

Summary Statistics for TRTN by AVAL
Controlling for STRTRD1

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	2.9188	0.0876
2	Row Mean Scores Differ	1	2.9188	0.0876
3	General Association	1	2.9188	0.0876

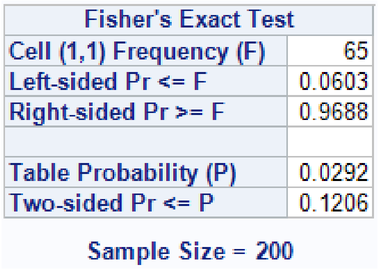
Figure 3 CMH Statistics Summary with Stratification on ORR

The SAS® PROC FREQ can also perform this as a non-stratificatied analysis per the code below. This is not feasible for our data, as we have shown the region has a confounding effect on treatment and response, however we wanted to show how this would work if there were no confounding. In that case, a chi-square test is appropriate:

```
ods listing close;
ods output FishersExact=<output dataset>;
proc freq data=<source dataset>;
   tables trt01pn*aval/chisq;
run;
ods listing;
```

In the non-stratification analysis, we wanted to observe the simplified association between treatment (variable TRT01PN) and outcome (variable AVAL). Considering cells may have fewer than 5 subjects, we selected the Fisher's exact test as the statistical method. The value of variable AVAL is 1 for the subjects with ORR and 0 without ORR.

Still using PROC FREQ, the option CHISQ requests chi-square tests and measurements. The tests include the Pearson chi-square, likelihood-ratio chi-square, and Mantel-Haenszel chi-square, as well as Fisher's exact test for 2x2 tables. Selecting ODS OUTPUT FishersExact will output a set of Fisher's exact test statistics in the indicated output dataset.



Fisher's Exact Test	
Cell (1,1) Frequency (F)	65
Left-sided Pr <= F	0.0603
Right-sided Pr >= F	0.9688
Table Probability (P)	0.0292
Two-sided Pr <= P	0.1206

Sample Size = 200

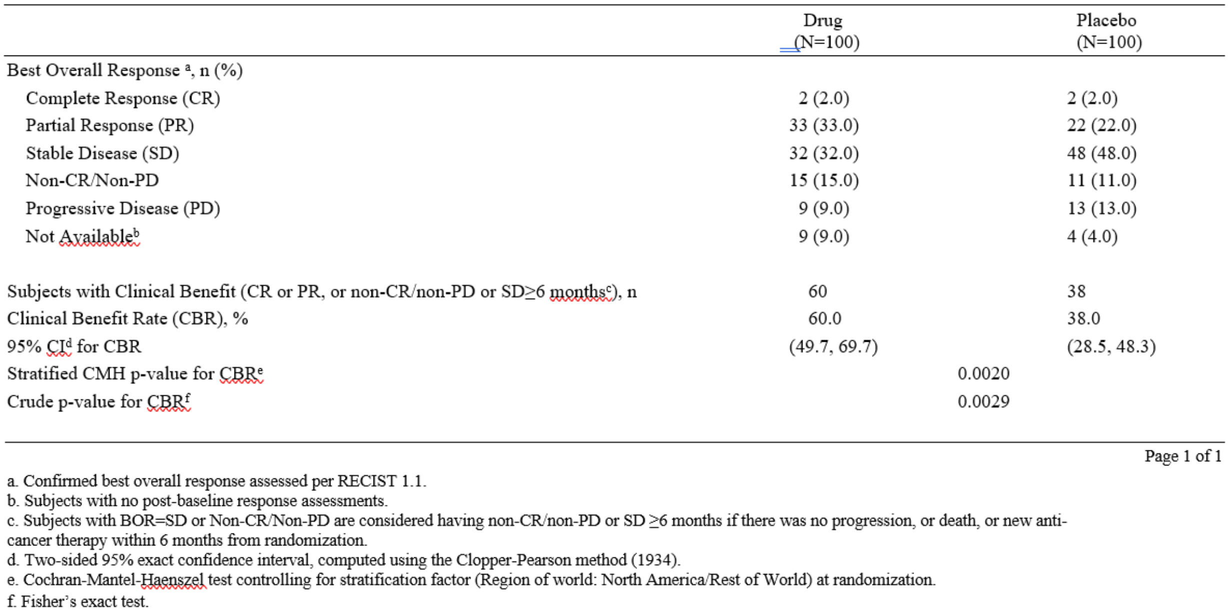
Figure 4: Fisher's Exact Test Statistics Summary Without Stratification on ORR

In Figure 2, the ORR point estimate in the treatment arm is 35% and comparatively higher than the placebo arm's 24%. However, the 95% CIs are overlapping (25.7% - 45.2% in treatment versus 16.0% - 33.6% in placebo), which indicates the ORR between the two arms is not significantly different. This is observed in the p-value as well, which - when considering regional adjustment - is 0.0889 which is more precise than the unadjusted p-value. Since we already learned region plays a confounding role in the model, we consider CMH p-value is more appropriate.

STRATIFIED CLINICAL BENEFIT RATE

In our example, clinical benefit is defined as achieving stable disease (SD) or non-CR/non-PD for ≥ 6 months or a best overall response of complete (CR) or partial response (PR) using RECIST 1.1.

Similar to the above, Figure 5 is shown based on the example data. Clinical benefit rates and the stratified p-value are calculated based on the FREQ procedure.

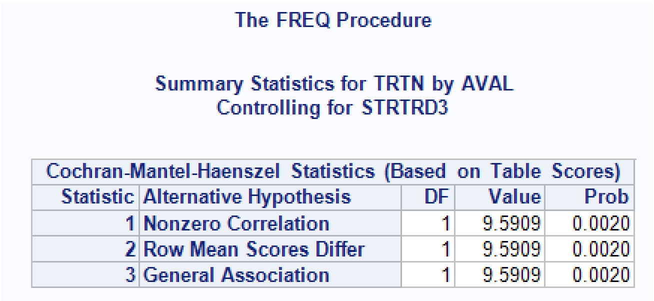


	Drug (N=100)		Placebo (N=100)
Best Overall Response [a], n (%)			
Complete Response (CR)	2 (2.0)		2 (2.0)
Partial Response (PR)	33 (33.0)		22 (22.0)
Stable Disease (SD)	32 (32.0)		48 (48.0)
Non-CR/Non-PD	15 (15.0)		11 (11.0)
Progressive Disease (PD)	9 (9.0)		13 (13.0)
Not Available[b]	9 (9.0)		4 (4.0)
Subjects with Clinical Benefit (CR or PR, or non-CR/non-PD or SD≥6 months[c]), n	60		38
Clinical Benefit Rate (CBR), %	60.0		38.0
95% CI[d] for CBR	(49.7, 69.7)		(28.5, 48.3)
Stratified CMH p-value for CBR[e]		0.0020	
Crude p-value for CBR[f]		0.0029	

Page 1 of 1

a. Confirmed best overall response assessed per RECIST 1.1.
b. Subjects with no post-baseline response assessments.
c. Subjects with BOR=SD or Non-CR/Non-PD are considered having non-CR/non-PD or SD ≥6 months if there was no progression, or death, or new anti-cancer therapy within 6 months from randomization.
d. Two-sided 95% exact confidence interval, computed using the Clopper-Pearson method (1934).
e. Cochran-Mantel-Haenszel test controlling for stratification factor (Region of world: North America/Rest of World) at randomization.
f. Fisher's exact test.

Figure 5. Example Output for the Clinical Benefit Rate



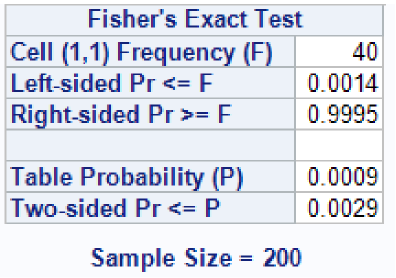
The FREQ Procedure

Summary Statistics for TRTN by AVAL
Controlling for STRTRD3

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	9.5909	0.0020
2	Row Mean Scores Differ	1	9.5909	0.0020
3	General Association	1	9.5909	0.0020

Figure 6. CMH Statistics Summary for CBR with Stratification



Fisher's Exact Test	
Cell (1,1) Frequency (F)	40
Left-sided Pr <= F	0.0014
Right-sided Pr >= F	0.9995
Table Probability (P)	0.0009
Two-sided Pr <= P	0.0029

Sample Size = 200

Figure 7. Fisher's Exact Test Statistics Summary for CBR Without Stratification

The CBR is significantly higher in the treatment group than in the placebo group: 60% (95% CI: 49.7% - 69.7%) on treatment whereas placebo shows 38% (95% CI: 28.5% - 48.3%). The p-value is pretty significant as shown: 0.0020 and 0.0029 with and without region adjustment, respectively. Since we know region stratification is needed, using CMH as our statistical method is more accurate for the stratification analysis.

CONCLUSION

CMH is a widely used statistical method to test the association between treatment and binary outcome while taking into account the stratification that controls for confounding factors. Since each stratum is homogeneous with regard to the confounder of interest, that helps to observe the association between exposure and outcome, adjusted by the effect of confounding. The SAS® FREQ procedure supports CMH analysis and can capture statistics using the ODS OUTPUT statement. In this paper, CMH concepts and ORR and CBR calculations with and without taking stratification consideration were discussed. In exploratory analysis, prior to deciding whether or not to stratify their analysis, a user should first observe if confounding plays a role by judiciously testing the data.

REFERENCES

U.S. Food and Drug Administration. 2018. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics

SAS Institute, Inc. 2014. SAS/STAT® 13.2 User's Guide: The FREQ Procedure. Cary, NC: SAS Institute Inc. Available at: https://support.sas.com/documentation/onlinedoc/stat/132/freq.pdf

Egeler, P.E. "Introduction to the Cochran-Mantel-Haenszel Test". Available at: https://cran.r-project.org/web/packages/samplesizeCMH/vignettes/samplesizeCMH-introduction.html

McDonald, J. H. "Cochran–Mantel–Haenszel Test for repeated rests of independence". Available at: http://www.biostathandbook.com/cmh.html

LaMorte, W. "The Cochran-Mantel-Haenszel Method". Available at: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_confounding-em/BS704-EP713_Confounding-EM7.html#headingtaglink_5

ACKNOWLEDGMENTS

We would like to acknowledge Grace An and Avani Kaja and Michiel Hagendoorn for their valuable suggestions to this paper.

CONTACT INFORMATION

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

Girish Kankipati
Seattle Genetics, Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
425-527-2140
gkankipati@seagen.com

Chia-Ling Ally Wu
Seattle Genetics, Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
awu@seagen.com