PharmaSUG 2020 - Paper SA-225

Principal Stratum strategy for handle intercurrent events: a causal estimand to avoid biased estimates - Case study and Power Simulations-

Andrea Nizzardo, Menarini Ricerche;
Giovanni Marino Merlo, Menarini Ricerche;
Simona Scartoni, Menarini Ricerche;

ABSTRACT

The ICH E9 (R1) addendum “Estimands and Sensitivity Analysis in Clinical Trial” emphasizes the necessity to quantify better treatment effects addressing the occurrence of intercurrent events that could lead to ambiguity in the estimates. One approach proposed in the addendum is the “Principal Stratum strategy,” where the target population for the analysis is a sub-population composed of patients free from intercurrent events. The main concern is that it is not possible to identify the stratum in advance, and the analysis is then not causal and liable to confounding. Moreover, the occurrence of intercurrent events is not predictable, and each subject is observed on one treatment only and could experience different intercurrent events on different treatments. Also the FDA Missing Data Working Group recommends the use of a causal estimand for the evaluation of primary interest endpoints.

A causal estimand based on principal stratum and a relative tipping point sensitive analysis is then proposed for a clinical superiority study. A case study is shown. An evaluation by simulations of the power obtained with this approach is also presented, comparing the ideal situation when all patients are free from intercurrent events and scenarios where different percentages of patients with intercurrent events are supposed.

INTRODUCTION

In clinical studies where two treatment arms are compared, commonly an experimental treatment and a reference therapy, the randomization process should ensure that the two treatment groups are comparable and homogeneous, i.e., balanced for those prognostic factors capable of influencing the clinical outcome of the patients. This condition is necessary to guarantee a causal relationship between the clinical efficacy observed at the end of the study and the treatment administered to the patients.

During the conduct of a clinical study, certain intercurrent events (e.g., death, intake of prohibited or rescue medications, adverse events leading to discontinuation, etc.) could occur after randomization and, therefore, cannot be considered independent of the assigned treatment. These events complicate the interpretation of the treatment effect.

The new ICH E9 (R1) addendum “Estimands and Sensitivity Analysis in Clinical Trial” emphasizes the necessity to estimate the treatment effect correctly by ad hoc estimands by incorporating all the intercurrent events occurred after the randomization and performing sensitivity analyses to confirm the results obtained. Missing Data Working Group of the U.S. Food and Drug Administration previously recommended the same approach for a correct evaluation of the treatment effect. The estimand definition and the methodology to incorporate the intercurrent events in the assessment of treatment effect needs to be clearly defined in the study protocol.

The ICH E9 (R1) addendum proposes several methods for the analysis. One of the most intuitive and easy to implement is the Principal Stratum approach in which all the patients with intercurrent events are excluded from the analysis. This method is similar to the Per Protocol analysis approach, and has the same limitations and issues. The events that determine the exclusion of patients from the analysis are events that occurred after the randomization process, and hence, cannot be considered independent of the treatment administered to the patients. The “causal property” of a suitable estimator is consequently affected, and the interpretability of the treatment effect is questionable.

Therefore, it is necessary to identify a method to modify the principal Stratum estimand and to maintain the causal relationship between the endpoints measured and the treatment taken by the patients.

METHODS

Lets consider a superiority trial analyzed with Principal Stratum approach where:

- T is the treatment (i.e., T_0=Reference, T_1=Test)
- Y is the outcome of interest
- S is the status indicating the occurrence of intercurrent events after the treatment intake (i.e., s=Occurred, $\bar{s}$=Not occurred)

The target population of the analysis based on Principal Stratum is composed of patients free from intercurrent events. The number of patients used for the analysis is, then, lower or equal to the one planned at the beginning of the study.

The bias of this approach is that the intercurrent events observed in a patient allocated to T_0 is not necessarily seen if the same patient had been treated with T_1. The estimand based on Principal Stratum is not causal.

To solve this issue, a causal principal stratification introduced by *Frangakis & Rubin's (2002)* and applied to the bioequivalence framework by *Lou et al. (2018)* is considered.

Let P be the stratum jointly identified by the occurrence of intercurrent events and the treatment:

$P = \{S_0, S_1\} = \{ss, s\bar{s}, \bar{s}s, \bar{s}\bar{s}\}$

where, for example, $s\bar{s}$ represents patients with intercurrent events if treated with T_0 and without intercurrent events, if treated with T_1.

The related causal estimand is the "Survivor Average Causal Effect" δ_{SACE} and it is defined as

$E(Y_1|P = \bar{s}\bar{s}) - E(Y_0|P = \bar{s}\bar{s})$, where the stratum $\bar{s}\bar{s}$ includes patients free from intercurrent events under both T_0 and T_1.

In a randomized superiority trial it is not possible to identify patients in $\bar{s}\bar{s}$. For this reason, a tipping point sensitivity analysis proposed by *Chiba & VanderWeele (2011)* is then used:

The $\hat{\delta}_{SACE}$ estimator is divided into two parts: the observed Principal Stratum estimand $\widehat{\delta_{PS}}$ that is defined as

$E(Y_1|P = \bar{s}) - E(Y_0|P = \bar{s})$ and the bias that represents the deviation from the true causal effect.

The Bias is defined as $\frac{\hat{\pi}_{\bar{s}s}}{\hat{p}_0}\beta_0 - \frac{\hat{p}_1 - \hat{p}_0 + \hat{\pi}_{\bar{s}s}}{\hat{p}_1}\beta_1$ where p_0 and p_1 represent the proportion of patients without intercurrent events if $T = 0$ and $T = 1$ respectively, and they are directly estimated from the observed data.

The other parameters are fixed but unknown:

- $\beta_0 = E(Y_0|S = \bar{s}s) - E(Y_0|S = \bar{s}\bar{s})$ that is the potential difference in outcome for patients treated with $T = 0$ between the stratum $\bar{s}s$ and $\bar{s}\bar{s}$.
- $\beta_1 = E(Y_1|S = s\bar{s}) - E(Y_1|S = \bar{s}\bar{s})$ that represents the potential difference in outcome in patients treated with $T = 1$ between the stratum $s\bar{s}$ and $\bar{s}\bar{s}$.
- $\pi_{\bar{s}s}$ that is the proportion of patients without intercurrent events only if treated with $T = 0$.

The tipping point sensitivity analysis is implemented by varying the unknown parameters. $\pi_{\bar{s}s}$ is bounded within the range $\pi_{\bar{s}s} \in [max(0, \hat{p}_0 - \hat{p}_1), min(\hat{p}_0, 1 - \hat{p}_1)]$, while β_0 and β_1 are free to vary within the clinically meaningful range usually derived from the observed data (e.g., quartiles, maximum, etc.).

In case of a high proportion of patients with IE, the $\widehat{\delta_{PS}}$ should be biased because of the reduced sample and the related power loss. However, *Lou et al. (2018)* demonstrated that the $\hat{\delta}_{SACE}$ estimator is unbiased whereas the $\widehat{\delta_{PS}}$ is biased unless there are equal β_0 and β_1. Additionally, they demonstrated that the power obtained by using the $\hat{\delta}_{SACE}$ estimator always increases if compared to the $\widehat{\delta_{PS}}$ power, except in very few peculiar cases.

To quantify the difference in power from the ideal situation where all patients are free from intercurrent events and the $\hat{\delta}_{SACE}$ power in case of intercurrent events, we have conducted a simulation study where different percentages of patients with intercurrent events are supposed.

We simulated three thousand studies with two treatment arms (i.e., reference and test) randomly assigned in a 1:1 ratio in which 304, 344, and 440 patients are considered (one thousand for each sample size). These numbers of patients are necessary to achieve a power of 80%, 85%, and 90% respectively, considering a Δ between the two treatment arms of 5, a standard deviation of 15.5, and an alpha level of 0.05 two-sided.

We then assigned to each patient a P stratum (i.e., $ss, s\bar{s}, \bar{s}s, \bar{s}\bar{s}$). When ss is assigned, it indicates that the intercurrent event occurs regardless of the treatment assigned and the patient is, consequently, always excluded from the analysis. While, on the contrary, in case of $\bar{s}\bar{s}$ the patient is always considered in the analysis since no intercurrent events occur. If $s\bar{s}$ or $\bar{s}s$ are assigned the patients will be excluded from the analysis only if the treatment assigned is reference or test respectively.

To assign the P stratum, four different scenarios have been considered:

1) The occurrence of intercurrent events is independent of the treatment (i.e., $s\bar{s}$ and $\bar{s}s$ never assigned).
2) The dependence between the intercurrent event and the treatment is equal in the two treatment arms (i.e., $s\bar{s}$ and $\bar{s}s$ assigned with the same probability).
3) No patients with intercurrent events if treated with test and without intercurrent events if treated with reference (the probability to assign $\bar{s}s$ is 0).
4) No patients with intercurrent events if treated with reference and without intercurrent events if treated with test (the probability to assign $s\bar{s}$ is 0).

All the patients assigned to the stratum $\bar{s}\bar{s}$ will be part of the principal stratum; conversely, 100% of the patients in ss will be discarded from the analysis. Regarding the strata $s\bar{s}$ or $\bar{s}s$, since the 1:1 ratio between the two treatments, 50% of the subjects will be excluded from each group. Practically speaking, the percentage of patients excluded from the principal stratum is obtained by adding the probability of being part of the stratum ss with half of the probability of being part of $s\bar{s}$ and $\bar{s}s$.

By varying the percentage of subjects assigned to the four strata, we can obtain, respecting the constraints specified for each scenario, different rates of patients without intercurrent events. To detect how the power of the SACE sensitivity estimator departs from the power value planned in the study protocol, we consider a percentage of missing subjects from 10% to 50%.

For the patients in the principal stratum, the outcome is assigned as follows:

- For the i^{th} subject assigned to the $\bar{s}\bar{s}$ stratum, we generate independent normal potential outcomes Y_1 and Y_0, depending on the treatment, as
$$Y_{1i} \sim N(60 + \Delta, 15.5^2)$$
$$Y_{0i} \sim N(60, 15.5^2)$$

- For the i^{th} subject assigned to the $s\bar{s}$ stratum (without intercurrent event only in the test group), we generate independent normal potential outcomes as
$$Y_{1i} \sim N(60 + \beta_1 + \Delta, 15.5^2)$$

- For the i^{th} subject assigned to the $\bar{s}s$ stratum (without intercurrent event only in the reference group), we generate independent normal potential outcomes as
$$Y_{0i} \sim N(60 + \beta_0, 15.5^2)$$

- For all other patients, who are not in the principal stratum, the outcome is not defined.

We consider in the simulation the following couples of (β_0, β_1) = (3, -3), (-3, -9), (-3, 3), (-3, 9), (-9, -3), (-9, -9), (-9, 3), (-9, 9), (3, -3), (3, -9), (3, 3), (3, 9), (9, -3), (9, -9), (9, 3), (9, 9); consequently, each study is replicated sixteen times for a total of 16000 sub-studies for each sample size.

The power of the SACE sensitivity estimator is obtained as the fraction of the 16000 sub-studies in which the null hypothesis of equality between treatment arms is rejected. In fact, since the value of Δ is set to 5 in the simulation, we are under H_1.

The model used to detect if the treatments are equal or not is an ANOVA model in which the dependent variable Y is the outcome, and the treatment arm is the fixed effect.

CASE STUDY

Let a new fixed combination of dexketoprofen trometamol (DKP.TRIS) and tramadol hydrochloride (TRAM.HCL), two analgesics with different mechanisms of action, be the new therapy to be tested for analgesic effect of moderate to severe acute pain. The primary objective of the study is the experimental evidence of superiority of the combined therapy vs. the monotherapies.

Rescue medication (RM) consisting of metamizole 500mg tablets is available on request during the treatment period. Besides, Paracetamol 500mg is also allowed to be used as an antipyretic agent when strictly necessary.

The primary efficacy evaluation is based on the patients' assessment of pain intensity at rest based on a 100 mm Visual Analogue Scale (VAS) ranging from "no pain" (i.e., VAS=0) to "worst pain imaginable" (i.e., VAS=100) recorded over 56 hours at different time-points.

The primary variable of interest, in testing the superiority of DKP.TRIS + TRAM.HCL versus DKP.TRIS and TRAM.HCL administered as single agents, is the average Sum of Pain Intensity Difference at rest at 56 hours (SPID56) calculated as the weighted sum of the PID values over 56 hours from the first treatment intake.

The formula for computing the SPID value is the following:

$SPID56 = \Sigma\,[T(i) - T(i-1)]\ x\ PID(i)$ for i = 1,…,32

Where:

- i represents the scheduled time point (i.e., t_{30m}, t_{1h}, t_{1h30m}, t_{2h}, t_{3h}, t_{4h}, and from t_{6h} to t_{56h} every 2 hours)
- T(i) is the scheduled time
- T(0)=0 (i.e. baseline)
- PID(i)= PID at time i= PI(0)- PI(i).

The following hypotheses are performed (μ denotes the respective mean):

H_0: μ(DKP.TRIS + TRAM.HCl) = μ(DKP.TRIS) *against* H_1: μ(DKP.TRIS + TRAM.HCl) ≠ μ(DKP.TRIS)

H_0: μ(DKP.TRIS + TRAM.HCl) = μ(TRAM.HCl) *against* H_1: μ(DKP.TRIS + TRAM.HCl) ≠ μ(TRAM.HCl)

The SPID56 is analyzed using an Analysis of Variance (ANOVA) model with treatment as a fixed factor and adjusted for the baseline pain intensity at rest.

A total number of 933 patients suffering from moderate to severe postoperative pain are randomized to one of the three treatment arms for the evaluation of analgesic effect. Among these patients, some needed to take the rescue medication during the treatment period.

Arm	***Treatment arm***	***Randomized Patients***	***Patients free from RM***
1	DKP.TRIS + TRAM.HCL	311	265 (85.2%)
2	DKP.TRIS	312	227 (73%)
3	TRAM.HCL	310	226 (73%)

Principal Stratum Estimand

Following the Principal Stratum approach, all patients taking the rescue medication are discarded from the analysis, which, consequently, is conducted considering a total of 718 patients.

The results below show that the combination therapy is superior to both the single agents (95% confidence interval does not contain 0).

δ_{PS}	LS means (SE)	95% CI
DKP-TRAM vs. DKP	-292.6 (84.7)	-459.0; -126.3
DKP-TRAM vs. TRAM	-223.5 (84.7)	-389.9; -57.2

However, the results obtained are questionable because the exclusion of patients from the analysis is due to intercurrent events occurred after randomization. The $\hat{\delta}_{SACE}$ estimator and the relative tipping point sensitivity analysis is then constructed.

SACE estimator and tipping point sensitivity analysis

To implement the SACE sensitivity estimator, the observed proportion of Principal Stratum subjects is firstly obtained for each treatment (i.e., 0.85 for the combination therapy, 0.73 for both the single agents).

Then, the values of the unknown parameters β_0, β_1 and $\pi_{\bar{s}s}$ are specified:

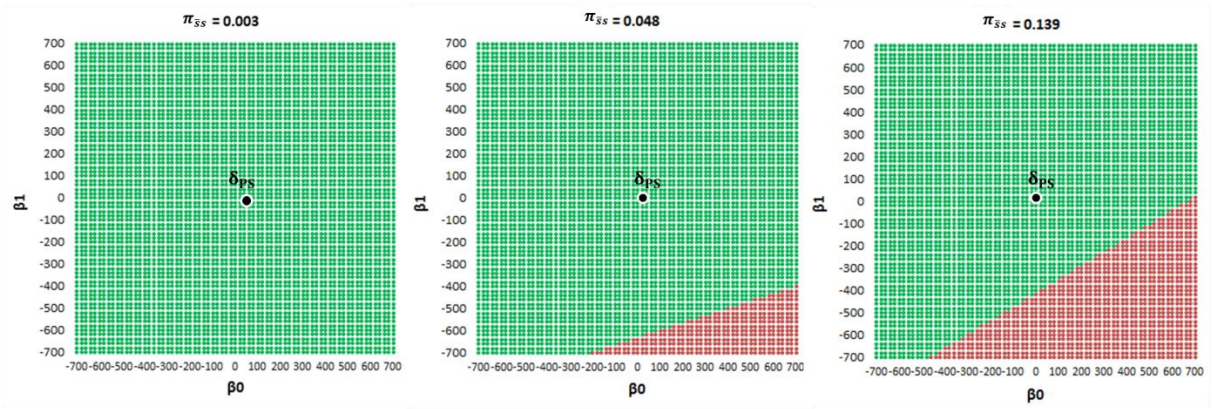
$\boldsymbol{\delta_{PS}}$	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\pi_{\bar{s}s}}$
DKP-TRAM vs. DKP	(-700;700) by 20	(-700;700) by 20	0;0.148
DKP-TRAM vs. TRAM			0;0.148

β_0 and β_1 are chosen due to their clinical significance and considering the observed median values.

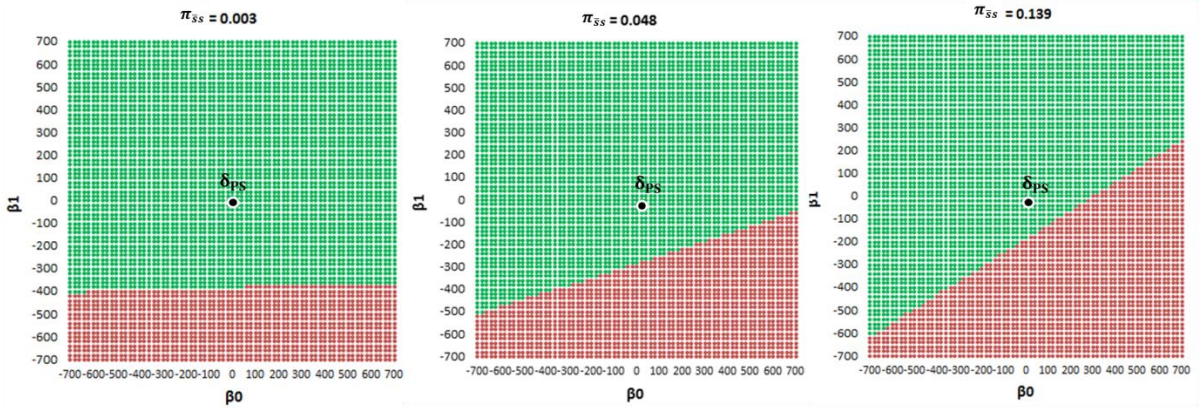
The values of $\pi_{\bar{s}s}$ are taken from the interval ($\pi_{\bar{s}s} \in [max(0, \hat{p}_0 - \hat{p}_1), min(\hat{p}_0, 1 - \hat{p}_1)]$). We chose three different values, i.e., 0.003, 0.048, and 0.139, that represent the extreme and the median values of the interval.

The results of the tipping point sensitivity analysis are shown below:

DKP-TRAM vs. DKP



DKP-TRAM vs. TRAM



In each block, one of the three chosen $\pi_{\bar{s}\bar{s}}$ proportion $\pi_{\bar{s}s}$ is fixed, and the superiority conclusion is plotted based on the sensitivity estimator for SACE corresponding to each paired value of β_0 and β_1 (5041 total combinations).

Below are the proportions of rejection for each value of $\pi_{\bar{s}s}$ for the contrast DKP-TRAM vs. DKP:

	Values of $\pi_{\bar{s}s}$			
Reject	**0.003**	**0.05**	**0.139**	
N	0 (0%)	376(7.46%)	1102 (21.86%)	1478 (9.77%)
Y	5041(100%)	4665(92.54%)	3939 (78.14%)	13645 (90.23%)
Total	5041	5041	5041	15123

Below are the results for the contrast DKP-TRAM vs. TRAM:

	Values of $\pi_{\bar{s}s}$			
Reject	**0.003**	**0.05**	**0.139**	
N	1165 (23.11%)	1521(30.17%)	1864 (36.98%)	4550 (30.09%)
Y	3876 (76.89%)	3520(69.83%)	3177 (63.02%)	10573 (69.91%)
Total	5041	5041	5041	15123

The combination therapy is superior to the single agents in the majority of sensitivity scenarios (i.e., the probability of superiority is at least > 0.75 for DKP-TRAM vs. DKP and > 0.60 for DKP-TRAM vs. TRAM).

Superiority conclusions obtained from the subjects in the principal stratum are, hence, robust.

POWER SIMULATION

The SACE sensitivity estimator is indeed an optimal method to maintain the causal property for a correct evaluation of the treatment effect. Nevertheless, the power associated with this approach is lower to the one that is planned in the study protocol, in an optimal situation where all the treated patients are free from intercurrent events.

To evaluate the difference between the power of $\hat{\delta}_{SACE}$ compared to the planned power, three sample sizes of 304, 344, and 404, with respective power of 0.80, 0.85 and 0.90, are considered. The loss in power using the SACE sensitivity estimator, considering different percentages of patients excluded from the analysis due to intercurrent events, is then calculated.

Each patient is assigned to the P stratum (i.e., $ss, s\bar{s}, \bar{s}s, \bar{s}\bar{s}$), identifying four criteria:

1) Strata $s\bar{s}$ and $\bar{s}s$ always empty.

   In this situation, the probability of having intercurrent events is independent of the assigned treatment. Consequently, the principal stratum approach does not violate the randomization, and it will approximately provide the same results as the SACE estimator. The missing percentages are obtained only by varying the probability of being ss (or $\bar{s}\bar{s}$)

2) Strata $s\bar{s}$ and $\bar{s}s$ assigned with the same probability.

   In this second criterion, the probability of intercurrent events is identical in both the treatment arms. For example, to obtain 40% of patients discarded from the principal strata ss is set equal to 5%, $\bar{s}\bar{s}$ is set equal to 25% and both $s\bar{s}$ and $\bar{s}s$ are set equal to 35%.

3) Probability to assign $\bar{s}s$ equal to 0.

   It is a situation in which there are no patients without IE in the reference group who would have IE if assigned to the test group. Conversely, there are plenty of patients without IE if assigned to the test group who would have IE in the comparison group. Each percentage of patients excluded from the PS is obtained by maximizing the difference between $\bar{s}s$ and $s\bar{s}$.

   For example, to get 30% of patients with IE, the strata are set as follow: ss equal to 5%, $\bar{s}\bar{s}$ equal to 45%, $\bar{s}s$ 0% and $s\bar{s}$ equal to 50%. To obtain, instead, 40% of patients excluded from the analysis, the strata are set as ss equal to 5%, $\bar{s}\bar{s}$ equal to 25%, $\bar{s}s$ 0% and $s\bar{s}$ equal to 70%.

4) Probability to assign $s\bar{s}$ equal to 0.

   The fourth scenario is obtained by reversing the probability of $s\bar{s}$ and $\bar{s}s$ used in the third criterion.

Note that, to generate a plausible situation, in the strata ss and $\bar{s}\bar{s}$ of each scenario, a percentage of patients at least equal to 5% is assigned, regardless of the sample size.

Below are the power loss from the ideal situation without intercurrent events by using the SACE sensitivity estimator.

80% power

N=304	Scenarios				
		First	Second	Third	Fourth
% of patients excluded from the analysis due to IE	10	0,044	0,047	0,052	0,051
	20	0,093	0,110	0,120	0,126
	30	0,154	0,172	0,220	0,229
	40	0,221	0,239	0,384	0,383
	50	0,298	0,303	0,663	0,629

85% power

N=344		Scenarios			
		First	Second	Third	Fourth
% of patients excluded from the analysis due to IE	10	0,046	0,045	0,048	0,049
	20	0,085	0,102	0,118	0,116
	30	0,151	0,167	0,218	0,220
	40	0,221	0,240	0,390	0,394
	50	0,295	0,304	0,704	0,663

90% power

N=404		Scenarios			
		First	Second	Third	Fourth
% of patients excluded from the analysis due to IE	10	0,031	0,037	0,039	0,038
	20	0,079	0,082	0,106	0,095
	30	0,133	0,152	0,198	0,207
	40	0,204	0,220	0,390	0,389
	50	0,276	0,290	0,731	0,695

When the level of missing information is moderate, i.e., 10% and 20%, the loss in power by using the SACE approach is similar in all the four supposed scenarios. Conversely, with higher levels of patients excluded from the principal stratum, the third and the fourth scenarios lead to a higher power loss compared to the first two cases.

The first and the second scenarios show similar results regardless of the percentage of patients without intercurrent events; the second one indicates slightly worse results, but the discrepancies are always below 2%. The third and the fourth scenarios show the same results for all percentages of excluded patients except for the last case (50%), where the third one seems to lead to more significant power loss than the fourth.

Note that the percentage of patients excluded from the principal stratum varies up to 50% because an adverse situation, in which almost all study patients have intercurrent events, is often implausible. In that unlikely case, we discourage the utilization of the principal stratum estimand.

This simulation study can be useful to forecast the power loss when using the causal version of the principal stratum estimand, i.e., the SACE.

Clinical information (e.g., the percentage of patients in each stratum), in addition to the usual information necessary for determining the sample size (e.g., Δ value), can be used together to maintain an acceptable level of study power when using the SACE approach.

From the three tables above, it is clear that, by increasing the expected study power, the SACE approach offers slight improvements. Comparing the 80% and 90% tables and considering, for example, the fourth scenario and 20% of excluded patients, the method approaches the ideal target of more than 3%.

In the code section of this paper we have presented a SAS macro which, based on the parameters chosen by the user, simulates thousands of studies to compute the power loss with respect to the one planned by using the PS and SACE approaches.

The SAS program allows a user to estimate the power loss caused by an approach based on a principal stratum in a two treatment arm study; consequently, the user is helped to choose an adequate sample size to prevent unpowered results.

CONCLUSION

This paper aims to examine the Principal Stratum approach, one of the methods suggested by EMA in the new ICH E9 (R1) addendum “Estimands and Sensitivity Analysis in Clinical Trial” to better quantify treatment effects addressing the occurrence of intercurrent events that could lead to ambiguity in the estimates.

This simple method is based on exclusion of patients with intercurrent events from the analysis, but it can be biased. Intercurrent events mostly occur after the randomization and, therefore, they cannot be considered independent of the assigned treatment. This situation can alter the comparability of the groups identified by the randomization process.

The causal version of the Principal Stratum (i.e., the SACE approach) proposed by *Chiba & VanderWeele (2011)*, is then recommended.

The application of this approach as sensitivity estimator in a superiority trial is presented in a case study where 25% of the patients experimented at least one intercurrent event.

The SACE confirmed the results obtained by using the principal stratum estimand and its causal property guarantee the robustness of study conclusions.

This method allows us to get unbiased estimates, nonetheless, it still considers fewer patients than the ones initially planned, thus leading to a potentially severe decrease of the study power.

For the reliability of results, it is essential to quantify the power loss in advance to plan an adequate number of patients enrolled.

The last section of the paper presents a macro program useful to forecast the power loss caused by the SACE method and, consequently, to choose a sample size in order to keep the study power into acceptable levels.

CODE

POWER SIMULATIONS

```
%MACRO PowerLoss(Nstudy=,sdev=,mu=,delta=,
B1MIN=,B1MAX=,B1INCR=,B0MIN=,B0MAX=,B0INCR=,
power=, alpha=,Pnn=,Pyn=,Pny=, Pyy=);

options nosource nonotes;
%let cnt=%sysfunc(countw(&power));

*find the sample size corresponding to each expected power;
proc power;
twosamplemeans test=diff
meandiff = &delta
stddev = &sdev
npergroup = .
alpha= &alpha
power = &power;
ods output output=want;
run;

*put the sample sizes into a macro variable;
Proc sql noprint;
Select (NPerGroup*2) into :N separated by " "
From want;  Quit;

%DO i=1 %TO &cnt.;
%LET NOK= %SCAN(&N.,&I.);
%LET powerOK= %SCAN(&power.,&i.);
%let blocksize = 2;

*block randomization;
data step1;
do study=1 to &Nstudy.;
do block = 1 to ceil(&Nok/&blocksize);
do item = 1 to &blocksize;
if item le &blocksize/2 then trt="test";else trt="ref";
rand = rand('UNIFORM');
output;
end;end;end;
run;

proc sort data = step1; by study block rand; run;

*strata assignment for each patient;
data Sample;
drop  prob:  x;
array prob [4] (&Pnn. &Pyn. &Pny. &Pyy.);
do study=1 to &Nstudy.;
do id = 1 to &Nok.;
x = rand("Table", of prob[*]);
if x=1 then strata="Pnn"; *always without intercurrent event;
if x=2 then strata="Pyn"; *without intercurrent event in test;
if x=3 then strata="Pny"; *without intercurrent event in ref;
if x=4 then strata="Pyy"; *always with intercurrent event;
output;
end; end;
run;

proc sort data=sample; by study id; run;
```

```
*principal stratum detection;
data step2;
merge step1(keep=study trt) sample;
if strata= "Pyy"  then PS="N";
if strata= "Pnn"  then PS="Y";
if strata= "Pny" and trt="ref" then PS="Y";
if strata= "Pny" and trt="test" then PS="N";
if strata= "Pyn" and trt="ref" then PS="N";
if strata= "Pyn" and trt="test" then PS="Y";
run;

*output simulation for each patient in PS;
DATA step3;set step2;
delta=&delta;mu=&mu;sdev=&sdev;
DO BETA1=&B1MIN TO &B1MAX BY &B1INCR;
DO BETA0=&B0MIN TO &B0MAX BY &B0INCR;
if strata="Pnn" and trt="test" then value = rand ('normal',mu+delta,sdev);
if strata="Pnn" and trt="ref" then value = rand ('normal',mu,sdev);
if strata="Pyn" and trt="test" then value = rand
('normal',mu+delta+beta1,sdev);
if strata="Pny" and trt="ref" then value = rand ('normal',mu+beta0,sdev);
if PS="N" then value=.;
OUTPUT;
end;end;
drop  mu sdev ;
run;

data step99;set step3;
group=cats(beta1, '/',beta0); value=round(value,0.001);
run;

ODS EXCLUDE all; ODS noRESULTS; ODS GRAPHICS off;
proc sort data=step99; by study group ;run;

*power of the principal stratum approach;
title 'anova ps';
proc mixed data=step99;
class  trt; model value=  trt /solution; by study group ;
lsmeans trt / diff=control("ref") cl;
ods output diffs=diffs ;
run;

data diffs1; set diffs;
IF STUDY=. then delete; if lower <= 0 then reject=0; else reject=1;
run;

proc sort data=step99;by study trt;run;

*finding p0 and p1 for the sace method;
proc freq data=step99;
table ps; by study trt;
ods output onewayfreqs=oneway (keep=study trt ps percent where=(ps="Y"));
run;

data oneway1 oneway2; set oneway;
percent=percent/100;
if trt="test" then output oneway1;
if trt="ref" then output oneway2;
run;

data diffs99;
```

```
merge diffs1 (drop=reject) oneway1 (keep=study percent
rename=(percent=percent_test) ) oneway2(keep=study percent
rename=(percent=percent_ref) );
by study;
if (percent_ref eq .) or (percent_test eq .) then delete;
run;

*SACE computation and power;
data sace; set diffs99;
beta1=input(substr(group, 1, index(group,"/")-1 ),8.);
beta0=input(substr(group, index(group,"/")+1, 6),8.);
UCLsace= upper+((&Pny./percent_ref)*beta0)-(((percent_test-
percent_ref+&pny.)/percent_test)*beta1);
LCLsace= lower+((&Pny./percent_ref)*beta0)-(((percent_test-
percent_ref+&pny.)/percent_test)*beta1);
if lclsace <= 0 then reject=0;
else reject=1;
run;

ODS EXCLUDE none; ODS RESULTS; ODS GRAPHICS ;
title "principal stratum power with N= &Nok.";
proc freq data=diffs1;
table reject*group /  norow nopercent ;
ods output crosstabfreqs=PS_&i.;
run;

title "sace sensitivity estimator power with N= &Nok.";
proc freq data=sace;
table reject*group  /  norow nopercent;
ods output crosstabfreqs=SACE_&i.;
run;

*putting together PS results and SACE results;
data final_&i.;
merge PS_&i.(rename=(Frequency=frequency_ps))
SACE_&i.(rename=(Frequency=frequency_sace));
N=&nok.;
if _type_ eq 10 and reject=1;
Ps_power=frequency_ps/(&nstudy.*16);
Sace_power=frequency_sace/(&nstudy.*16);
diff=sace_power-ps_power;
keep N frequency_sace frequency_ps sace_power ps_power;
run;

proc sort data=final_&i;by n;run;

*computation of the power loss vs. expected power with PS and SACE;
data loss_power_&i;SET  final_&i  ;
length Missing_percentage $8.;
PS_power=round(ps_power,.001);
Sace_power=round(Sace_power,.001); Expected_power=(&powerok./100);
PS_power_loss= round(expected_power - PS_power,.001);
Sace_power_loss= round(expected_power - sace_power,.001);
drop frequency_ps frequency_sace;
Missing_percentage= cats(%sysevalf((&pyy+&pyn*0.5+&pny*0.5)*100),"%");
run;
%end;

data loss_power;set loss_power_:; run;

proc datasets  lib= work    memtype=data;
```

```
delete diffs: final_: loss_power_: oneway: ps_: sace: sample step: want
;run;quit;

options source notes;
%MEND;

/*Specify:
-the number of study that will be simulated for each power level and for
each combination of beta values (1);
-the expected standard deviation of your data (2);
-the mean of the reference treatment (3);
-the expected difference between the test treatment and the reference
treatment (4);
-the minimum, the maximum and the increment value of Beta1 (5,6,7);
-the minimum, the maximum and the increment value of Beta0 (8,9,10);
-the desired power levels (11);
-the alpha value (12);
-the probability assigned to each stratum (13,14,15,16).

Note1: it is possible to specify more than one power level;
Note2: when specifying the expected power level (11), use the notation .xx
(e.g., 80% -> .80);
Note3: to specify only one value of beta1, set B1MIN equal to B1MAX and
B1INCR equal to 0(same for beta0);
Note4: the strata are the following:
       Pnn: subjects without intercurrent event regardless of the assigned
treatment;
       Pyn: subjects without intercurrent event only if test treatment is
assigned;
       Pny: subjects without intercurrent event only if reference treatment
is assigned;
       Pyy: subjects always with intercurrent event regardless of the
assigned treatment.
Note5: The missing percentage is obtained in this way:
(Pyy+(Pyn*0.5)+(Pny*0.5))*100;
*/

%PowerLoss(
/*1*/ Nstudy=500,
/*2*/ sdev=10,
/*3*/ mu=60,
/*4*/ delta=7,
/*5*/ B1MIN= -9, /*6*/ B1MAX=9, /*7*/  B1INCR=6,
/*8*/ B0MIN= -9, /*9*/ B0MAX=9, /*10*/ B0INCR=6,
/*11*/power= .80 .90,
/*12*/alpha= 0.05,
/*13*/Pnn=0.6, /*14*/Pyn=0.3, /*15*/Pny=0.1, /*16*/Pyy=0.0);
```

REFERENCES

- EMA/CHMP/ICH/436221/2017. ICH E9 (R1).
- Frangakis, C. E., Rubin, D.B. (2002).Principal stratification in causal inference.Biometrics 58:21-29.
- Chiba , Y. VanderWeele, T.J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. American journal of epidemiology 173(7):745-751.
- Lou, Y., Jones, M., Sun, W. (2018): Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: non compliance and missing data. Journal of biopharmaceutical statistics.
- Michele M. Burlew (2014), SAS macro programming made easy, Third Edition, SAS institute.