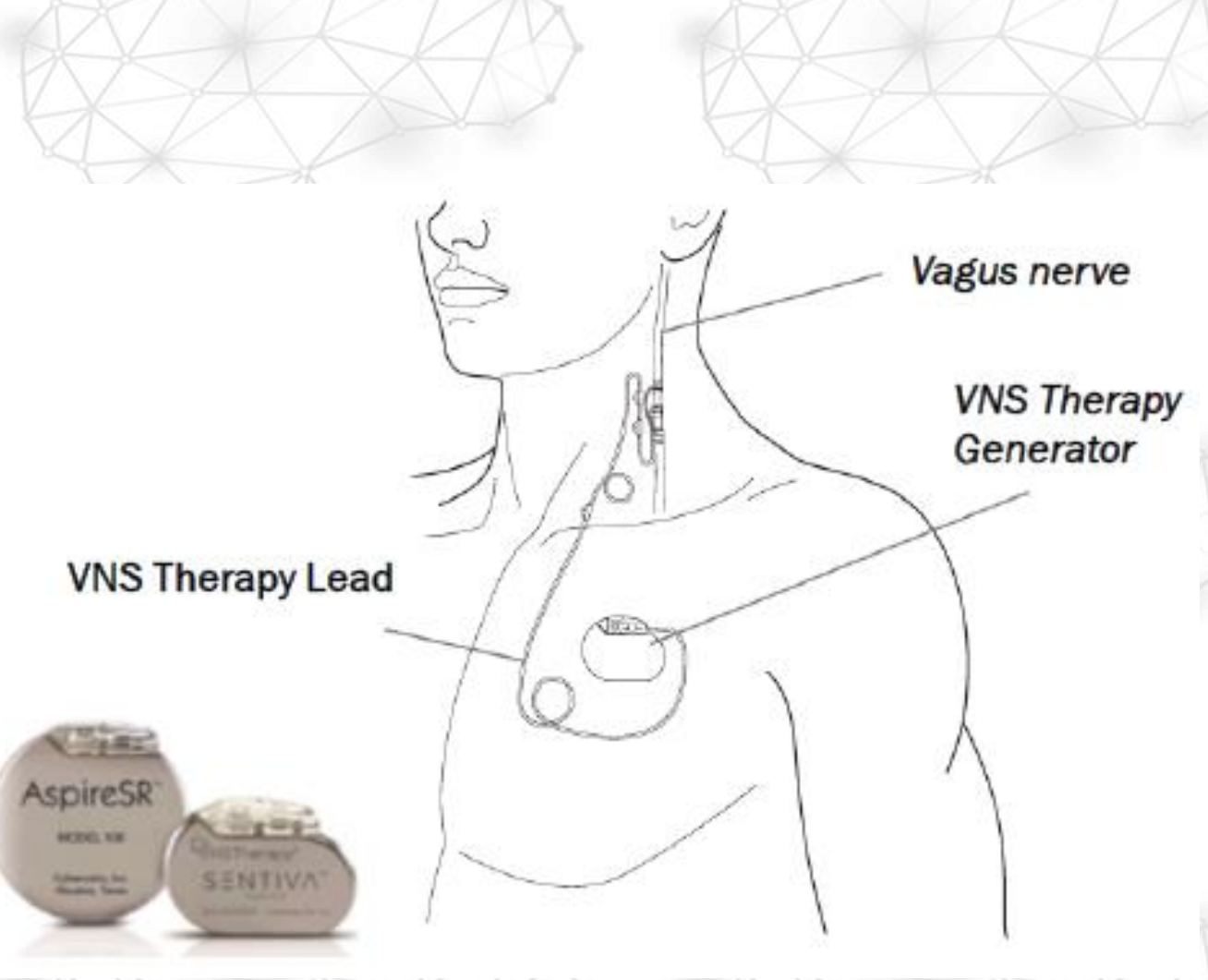


BACKGROUND



Picture 1. Vagus Nerve Stimulation Therapy

The Vagus Nerve Stimulation (VNS) Therapy (manufactured by LivaNova), is an approved adjunctive treatment delivered by an implantable device to treat epilepsy by reducing the frequency of seizures.

We explored a potential clinical trial design to evaluate the effectiveness, safety, and healthcare utilization outcomes of VNS Therapy for the treatment of drug-resistant epilepsy against the current pharmaceutical Standard of Care (SoC). The design described is a long-term prospective, multicenter, parallel-group, observational study.

The primary endpoint is the maximum duration of continued, monthly response to treatment. The treatment choice is decided by the physician and the patient prior to receiving the new treatment. From a statistical point of view, the issues are: a) Duration of response rarely follows a common distribution; and b)Accounting for baseline differences in the non-randomized treatment groups.

In the PP16 poster of PhUSE EU Connect 2022, we analyzed data with a non-parametric approach, using the overlap-weights technique for minimizing the differences between treatment groups, finding the best way of making inferences from them and comparing this approach with parametric ones.

By simulating different solutions based on a true study design, aimed at showing the difference between VNS Therapy against SoC, we discovered that the best way for doing inference in that context was calculating confidence intervals of the weighted difference in mean of the duration of maximum response with a bootstrapping non-parametric method. All other methods studied brought a Type I error inflation.

The goal of the current analysis is instead to understand how to make inferences in a situation where an Individual Patient Data (IPD) population is compared with results from the literature. The Matching-Adjusted Indirect Comparison (MAIC) method developed by Signorovitch et al [1] can reduce the differences in the baseline characteristics by applying a weight to IPDs. But how can we make inferences from these weighted data?

METHODS

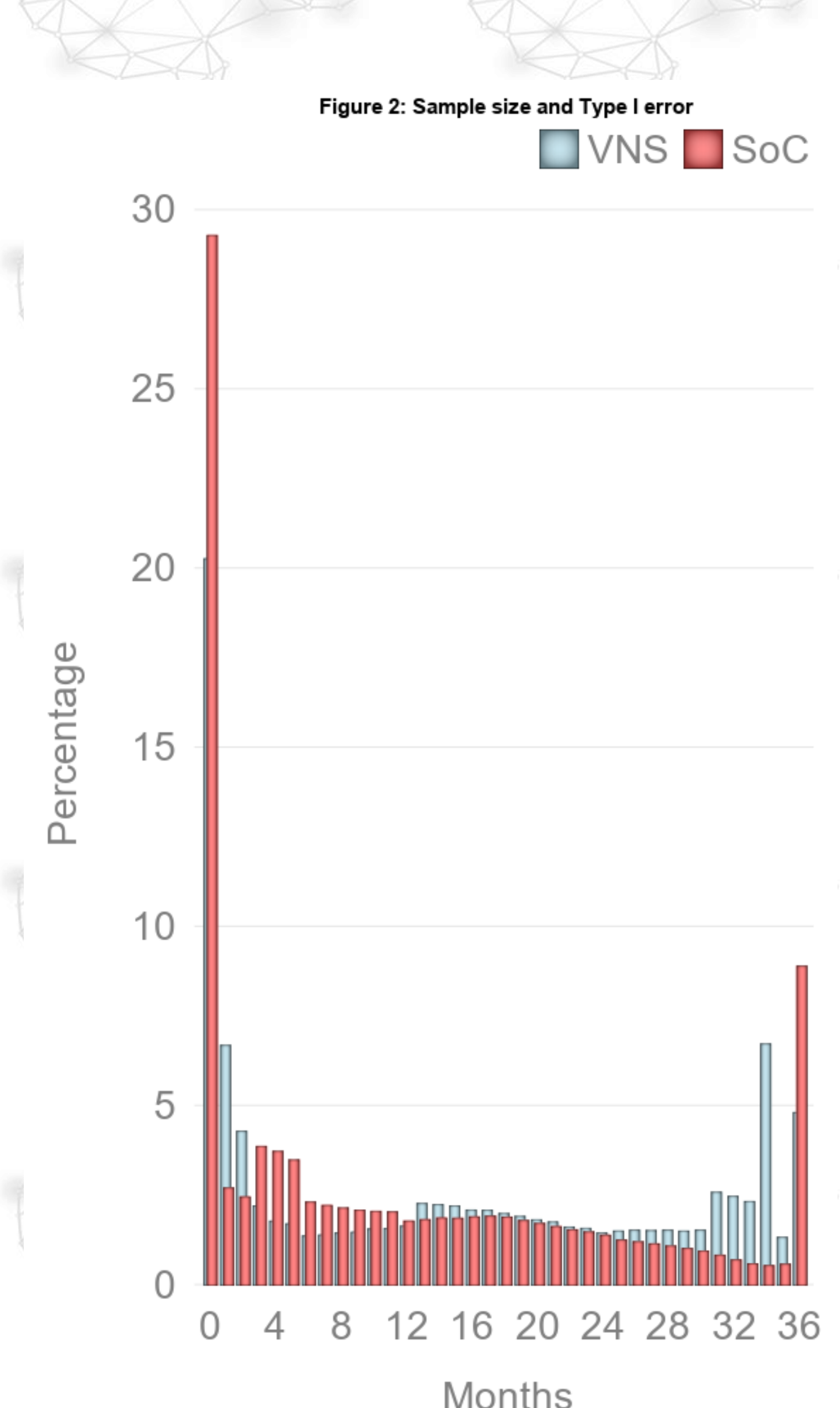


Figure 1. The maximum duration of response to treatment

Simulated population. Responses to therapies (VNS and SoC) have been simulated on a monthly basis for 1 000 000 subjects each, based on literature data. Each monthly response depends on baseline characteristics and the responses of the subject in the 3 previous months.

A total of 36 months has been simulated and the maximum duration of continued response to treatment is the primary endpoint. The distributions of these durations in the 2 populations are shown in **Figure 1**. The baseline characteristics of the 2 populations have been simulated based on input parameters enlisted in **Table 1**.

Table 1. Design parameters

Parameter	VNS	SoC
Cognitive Impairment /Learning Disability	60%	40%
Number of failed SoCs	Mean 6.8, SD 3.06 Min 3, Max 9	Mean 8.1, SD 4.1 Min 3, Max 9
Duration of epilepsy (years of Age minus Age at Onset)	Mean 14.4, SD 12.74 Min 0, Max 20	Mean 24.9, SD 13.4 Min 0, Max 20
Age	Mean 23.9, SD 16.55, Min 4, Max 70	Mean 36, SD 18, Min 4, Max 70
Seizure Type	Focal 70%, Generalized 30%	Focal 80%, Generalized 20%
Baseline Seizure Rate per Month	[2, 10): 28% [10, 100): 50% [100, 1000): 22%	[2, 10): 28% [10, 100): 50% [100, 1000): 22%

SAMPLE SIZE CALCULATION

Non-parametric, confidence interval calculated using bootstrapping. How many subjects are needed to get a significant superiority of VNS treatment in the maximum duration of continued response to treatment, considering a type I error level of 0.025 and a power of 0.80?

- From the VNS and SoC populations, a number of subjects equal to V and S respectively have been randomly extracted without replacement. This step has created 10 000 datasets for VNS subjects named VNS00001-VNS10000, as well as 10 000 datasets for SoC named SOC00001-SOC10000.
- For each SOC0001-SOC10000 dataset, the mean values of the 6 parameters reported in the Methods section have been calculated. For each VNS00001-VNS10000 dataset, the algorithm developed by Signorovitch et al. for calculating the balancing weights, has been applied to let the mean of each baseline parameter of dataset VNS00001 be equal to those of SOC00001, as well as for each pair (VNS0002-SOC0002 ... VNS10000-SOC10000).
- For each pair of datasets VNS0001-SOC0001, the weighted difference in the mean of the duration of maximum response in the two treatment groups has been calculated, together with its 95% confidence intervals calculated with the bootstrapping method (randomly extracting the subjects with replacement from VNS datasets, keeping a unique mean for the SOC ones).
- If the nth confidence interval did not include 0 in its values, the nth simulation is considered a success.

The power of the analysis with V VNS subjects and S SoC subjects has been calculated by dividing the number of successes by the 10 000 simulations done.

Steps 1-4 have been repeated with increasing V and S numbers until the first value gets a power ≥ 80%. The ratio V:S has been set to 1:2.

Parametric, confidence intervals were calculated using normal assumptions. Steps 1, 2, and 4 are in common with the first method described.

- For each pair of datasets VNS0001-SOC0001, VNS0002-SOC0002 ... VNS10000-SOC10000, the weighted difference in the mean of the duration of maximum response in the two treatment groups have been calculated and the 95% confidence intervals have been calculated using the following formula:

$$CI = Mean \pm z \sqrt{wSE_{VNS}^2 + wSE_{SOC}^2}$$

Parametric, confidence intervals were calculated using normal assumptions and rescaled weights.

The steps are the same as those in the parametric method described above, but the weights assigned to each subjects have been rescaled so that their sum equals the original number of subjects in the VNS sample.

RESULTS

Assessing the reliability of the methods. To assess if the used methods are reliable, the Type I error inflation has been evaluated. Based on the sample sizes found in the Sample Size Calculation section, the power is calculated using a population dataset including both the VNS and SoC ones. The expected power should be close to the 0.025 chosen as alpha for sample size calculation (**Figure 2**).

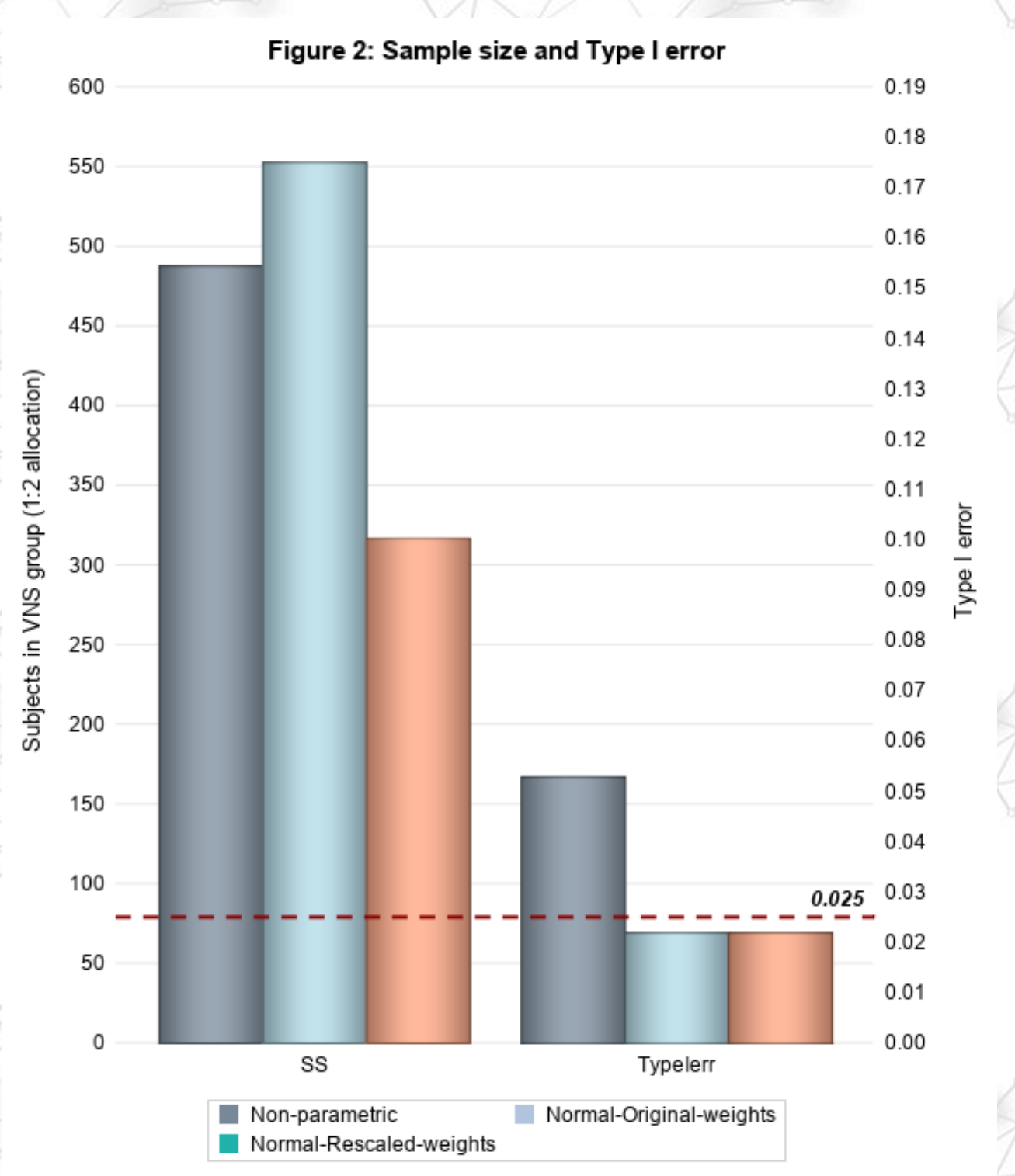


Figure 2. Sample size and Type I error

Results: Conversely to the past year simulation on 2 IPD data sources, the non-parametric method with bootstrap confidence intervals is inflating the Type I error (0.053). The two parametric calculations perform better, both with a Type I error of 0.022. As a best choice in this case, the inference considering rescaled weights is more powerful than inferencing with original weights, requiring a lower sample size (317 vs 553) to get the same power, preserving the type I error.

CONCLUSION

The results obtained in this analysis are not generalizable to other scenarios but are useful for understanding what is the correct inference method to be used in this analysis. The distribution of the endpoint is clearly not following any known distribution and the results can be influenced by this fact. The suggestion is to repeat this exercise ahead of each analysis to find the best inference method for that situation.

REFERENCES

[1] Signorovitch J. E. (2012). Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health*, 15(6), 940-947.