Evaluation of the Fill-it-up-design to use historical control data in randomized clinical trials with two arm parallel group design

S. Wied[1], M. Posch[2], R.-D. Hilgers[1]

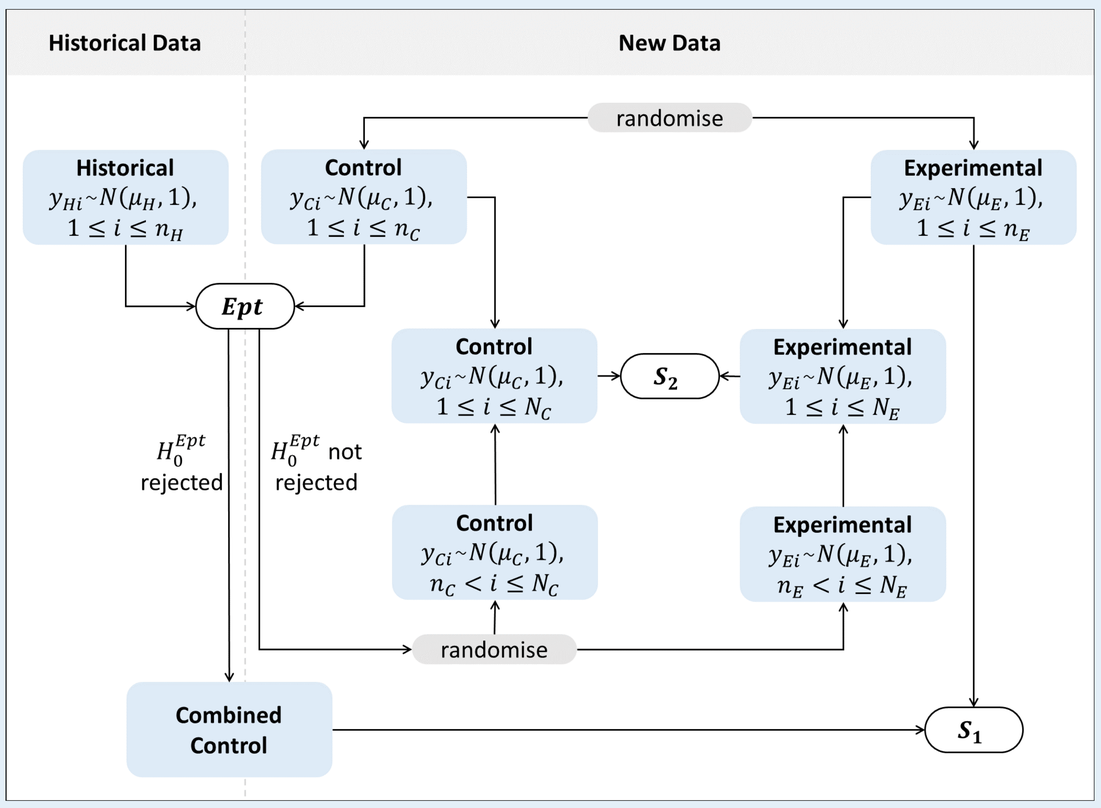
[1]Institute of Medical Statistics, RWTH Aachen University, Aachen, Germany

[2] Medical University of Vienna, Centre for Medical Data Science, Institute of Medical Statistic, Vienna, Austria

Introduction

The most appropriate method to assess the effect of an intervention in clinical research is to conduct a randomised controlled trial (RCT). In some areas, for example where small population groups are involved, conducting a RCT is challenging. With the help of historical controls, the existing information can be utilized to support the new study design, but of course, inclusion also carries the risk of bias in the study results. The objective of the presented Fill-it-up-design (FIU) is to demonstrate equivalence of historical and randomized controls in order to obtain unbiased estimates of treatment effect.

The Fill-it-up-design



Statistical Tests & Hypotheses:

$$(Ept): \quad H_0^{Ept}: |\mu_C - \mu_H| \ge \Delta \quad \text{vs.} \quad H_1^{Ept}: |\mu_C - \mu_H| < \Delta$$

$$(S1): \quad H_0^{S_1}: \mu_E - [\omega\mu_H + (1-\omega)\mu_C] \le 0 \quad \text{vs.}$$

$$H_1^{S_1}: \mu_E - [\omega\mu_H + (1-\omega)\mu_C] > 0,$$

where $\omega^* = n_H/(n_H + n_C)$ is the optimal weight.

$$(S2): \quad H_0^{S_2}: \mu_E - \mu_C \le 0 \quad \text{vs.} \quad H_1^{S_2}: \mu_E - \mu_C > 0$$

Decision function of the FIU can be written as

$$\psi_{FIU} = \max\left(\varphi_{Ept}^{\alpha_{Ept}} \varphi_{S_1}^{\alpha_{S_1}}, (1 - \varphi_{Ept}^{\alpha_{Ept}}) \varphi_{S_2}^{\alpha_{S_2}}\right).$$

Sample Size

Let

$$N_E = n_E + n_E', \; N_C = n_C' + n_C, \; \gamma N_E = n_E \text{ and } \gamma N_C = n_C,$$

where $(S1)$ is conducted with $n_E + n_C + n_H$ and $(S2)$ with $N_E + N_C = n_E + n_E' + n_C + n_C'$.

⇒ For equal type I and type II error probabilities $\alpha_{S_1} = \alpha_{S_2}$, $\beta_{S_1} = \beta_{S_2}$ as well as balanced sample sizes $N_E = N_C = N$ this results in

$$\gamma = \frac{1}{2N}\left(N - n_H + \sqrt{N^2 + n_H^2}\right).$$

- Total sample size of the FIU $N_{FIU} = N_E + N_C$ (no equivalence)
- Total number of patients recruited in the first step γN_{FIU}.

Familywise Error Rate

The family wise error rate (FWER) of the procedure should satisfy

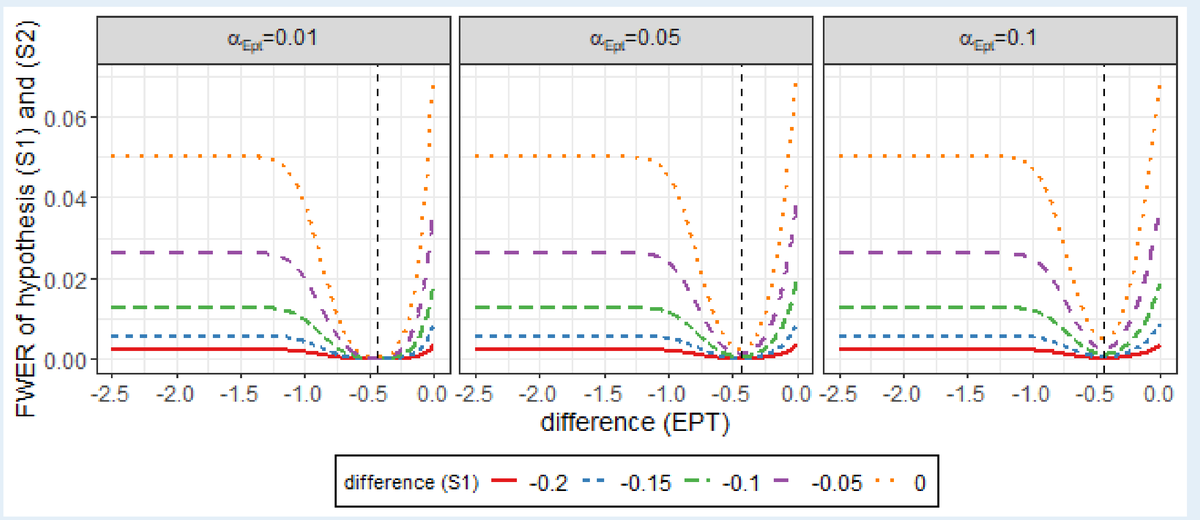
$$E(\psi_{FIU}) \le 0.05$$

and can be obtained by

$$E(\psi_{FIU}) = E\left(\varphi_{Ept}^{\alpha_{Ept}} \varphi_{S_1}^{\alpha_{S_1}}\right) + E\left(\left(1 - \varphi_{Ept}^{\alpha_{Ept}}\right) \varphi_{S_2}^{\alpha_{S_2}}\right).$$

Results

Under the respective null hypotheses H_0^{Ept} and $H_0^{S_1}$ (i.e. lines left to the dashed vertical line), the FWER is maximized for larger differences between the expected response of historical and current controls (difference (Ept)) together with a weighted treatment effect (difference (S_1)) of zero. Under the respective null hypotheses H_0^{Ept} and $H_0^{S_1}$ the maximum FWER is kept at 5%. Here the FWER is examined for a medium effect size $\delta = 0.5$ with $n_H = 500$ historical controls and an equivalence margin of $\Delta = 0.44$. Similar results are obtained for small and large effect sizes [4].



The Fill-it-up-design in practice

Between September 15, 2010, and November 21, 2013, 605 patients were enrolled in the prospective international registry investigating the natural history of Friedreich's ataxia [3]. The Scale for the Assessment and Rating of Ataxia (SARA) serves as a primary endpoint variable and was followed up yearly. In a fictive clinical trial with a two arm parallel group design, a new treatment should be compared to standard of care (SOC) with respect to the difference in mean SARA score two years after enrollment. To compare our method with the existing ones, we consider the MAP of [1]. For the application we use the following setting:

Parameter		Value
n_H	Sample size of historical controls	500
δ	Effect size	0.275
$1-\beta_{S_1} = 1-\beta_{S_2}$	Power of (S_1) and (S_2)	0.8
$\alpha_{S_1} = \alpha_{S_2}$	Significance Level of (S_1) and (S_2)	0.05
α_{Ept}	Significance Level of (Ept)	$\{0.01, 0.05, 0.1\}$
Δ	Equivalence Margin	$\left[\sqrt{\frac{1}{n_H} + \frac{1}{n_C}} \cdot z_{1-\alpha_{Ept}}, \delta\right)$

- Sample Size of FIU: $N_{FIU} = 328$ and $\gamma N_{FIU} = 192$.
- Sample Size of fixed sample design: 328 (S_2 only)

A simulation study with two different scenarios is carried out (50k simulation replications):

I (Minimum Case) → $\mu_C = 0$, $\mu_E = -\omega\Delta$, $\mu_H = -\Delta$

II (Maximum Case) → $\mu_C = 0$, $\mu_E = 0$, $\mu_H = 3\Delta$

			FIU	MAP	
Scenario	Δ	α_{Ept}	FWER	TIE (full-sample)	TIE (sub-sample)
I	0.27	0.01	0.0002	0.0002	0.0013
	0.22	0.05	0.0015	0.0007	0.0029
	0.19	0.1	0.0033	0.0015	0.0043
II	0.27	0.01	0.0519	0.0461	0.0447
	0.22	0.05	0.0519	0.0454	0.0439
	0.19	0.1	0.0519	0.0453	0.0437

The FIU can potentially save patients. However, one has to accept a small inflation of the error compared to the MAP approach.

Conclusion

- FIU is an easy applicable and comprehensive method for including historical control data.
- A robust prior belief is essential for its use.
- Can be seen as a way out in exceptional situations where a hybrid design is considered necessary
- We recommend to follow the six criteria of Pocock [2] to ensure the safe use of historical controls

References

[1] B. Neuenschwander et al.: Summarizing Historical Information on Controls in Clinical Trials. *Clinical Trials*, 7(1), 5-18, 2010.

[2] S. J. Pocock: The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29(3):175-188, 1976.

[3] K. Reetz et al.: Progression characteristics of the european friedreich's ataxia consortium for translational studies (efacts): a 2 year cohort study. *Lancet Neurol*, 15:1346-1354, 2016.

[4] S. Wied et al.: Evaluation of the Fill-it-up-design to use historical control data in randomized clinical trials with two arm parallel group design. *BMC Med Res Methodol* 24(197), 2024.

The IDeAl project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no. 602552.

This work is supported by the funding from the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP no. 825575.