Power Analysis for Testing Treatment-Biomarker Interaction in Two-Phase Design

Poulami Maitra
Research Statistics, GlaxoSmithKline

In collaboration with
Jianwen Cai & Donglin Zeng
Department of Biostatistics, University of North Carolina at Chapel Hill
Xiaofei Wang
Department of Biostatistics and Bioinformatics, Duke University

October 4, 2019

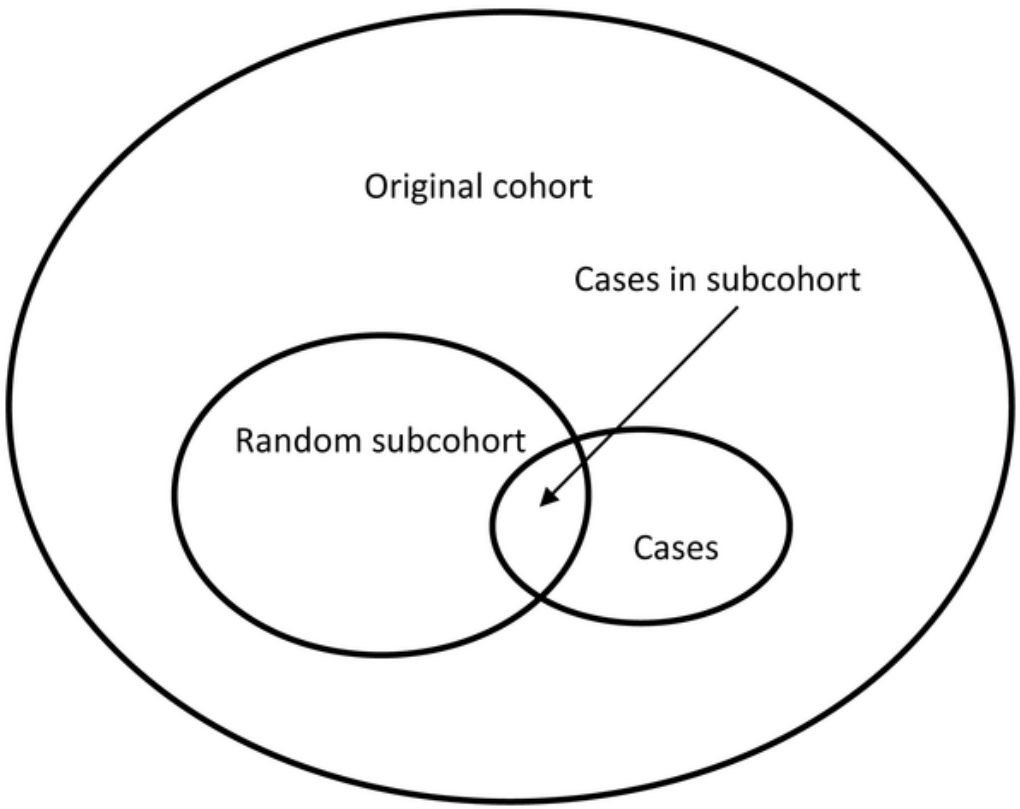
Introduction

Motivation

- Discovering new biomarkers to predict the best treatment for a patient.
- Finding new biomarkers is expensive.
- Large amounts of data are available after Phase III trials are completed.
  - Frozen blood/biosamples from patients.
- The advantages of a two-phase design:
  - The data collection time will be short.
  - Sampling of the second phase utilizes information from the first phase.
  - Expensive biospecimen can be collected.

Case-Cohort Studies for Failure Time Data

- Collecting Covariate and follow-up information constitutes majority of cost.
- Prentice (1986) proposed case-cohort design as a cost-effective alternative.



Existing Literature

- Two-phase design have been extensively studied in time to event analysis.
- Another aspect is the design of such studies.
- Cai and Zeng (2004) and Cai and Zeng(2007) proposed simple formula to calculate the power for the main effect under case-cohort studies and bounds under generalized case-cohort design, respectively.
- We develop power/sample size formula for testing the interaction between a treatment and an expensive biomarker.

Method

Method

- Two biomarker groups (0,1) and two treatment groups (0,1) : n_{jk} (j, k = 0, 1) individuals in each group.
- $T^*_{i,jk}$ be the potential failure time; $C_{i,jk}$ is the censoring time for individual i in the treatment group k and the biomarker group j.
- $T_{ijk} = min(T^*_{i,jk}, C_{i,jk})$ is the observed time.
- The hazard is defined as:

$$\lambda_{j,k}(t) = e^{\beta_j k}\lambda_j(t) \quad j,k = 0,1.$$

- The null hypothesis:

$$H_0 : \frac{\lambda_{11}(t)}{\lambda_{10}(t)} = \frac{\lambda_{01}(t)}{\lambda_{00}(t)} \forall t \Rightarrow \beta_1 = \beta_0.$$

Method

- $$\tilde{U}_j(\beta_j, 1) = \sum_{i=1}^{n_{j1}} \frac{\tilde{Y}_{j0}(T_{ij1})\Delta_{ij1} W(T_{ij1})}{\tilde{Y}_{j0}(T_{ij1}) + e^{\beta_j}\tilde{Y}_{j1}(T_{ij1})} - \sum_{i=1}^{n_{j0}} \frac{e^{\beta_j}\tilde{Y}_{j1}(T_{ij0})\Delta_{ij0} W(T_{ij0})}{\tilde{Y}_{j0}(T_{ij0}) + e^{\beta_j}\tilde{Y}_{j1}(T_{ij0})} \quad \forall j = 0, 1.$$
- $\hat{\beta}_j$ is the solution to $\tilde{U}_j(\beta_j, 1) = 0$.
- σ_j and δ_j are estimated from the variance estimators of $\hat{\beta}_j$.
- The proposed test is:

$$TS_n = n^{1/2} \frac{\{\hat{\beta}_1 - \hat{\beta}_0\}}{\sqrt{\sum_j \frac{n}{n_j}\left(\hat{\sigma}_j^{-2} + \hat{\sigma}_j^{-4}\hat{\delta}_j\right)}}. \quad (1)$$

Power formula

- We consider the case for w(t) = 1. We also assume the following :
  (a) The censoring distributions are the same in the two treatment groups for each of the biomarkers.
  (b) The proportion of failures is small in the full cohort.
  (c) No ties of failures are observed.

Power formula

Theorem

For the alternative hypothesis, $H_A : \beta_1 - \beta_0 = \theta > 0$, $\theta = O(\tilde{n}^{-1/2})$ (where $\tilde{n} = \sum_j \tilde{n}_j$ and $\tilde{n}_j$ are of the same order), the power of the test statistic TS_n can be approximated by

$$\Phi \left[\sqrt{n} \left(\beta_1 - \beta_0 \right) \left\{ \sum_{j=0}^{1} r_j^{-1} \left(\sigma_j^{-2} + \sigma_j^{-4} \delta_j \right) \right\}^{-1/2} - Z_{1-\alpha} \right] . \tag{2}$$

Power formula

Approximation 1

Assuming $\frac{m_{*j1}}{n_{j1}} \approx (1 - p_D^{j1})$ and $\frac{m_{*j0}}{n_{j0}} \approx (1 - p_D^{j0})$, the power of the test statistic TS_n can be approximated by

$$\Phi \left[\sqrt{n} (\beta_1 - \beta_0) \left\{ \sum_{j=0}^{1} r_j^{-1} \frac{1}{p_j(1-p_j) \left[e^{2\beta_j} p_j (1 - p_D^{j1})^2 p_D^{j0} + (1-p_j)(1-p_D^{j0})^2 p_D^{j1} \right]} \right.$$
$$\times \left(\left((1-p_j)(1-p_D^{j0}) + e^{\beta_j} p_j (1 - p_D^{j1}) \right)^2 \right.$$
$$\left. \left. + \frac{\left[e^{2\beta_j} (1-\psi_j)(1-p_D^{j1})(1-p_D^{j0}) \right] \left((1-p_j) p_D^{j0} + p_j p_D^{j1} \right)^2}{\psi_j \left(e^{2\beta_j} p_j (1-p_D^{j1})^2 p_D^{j0} + (1-p_j)(1-p_D^{j0})^2 p_D^{j1} \right)} \right) \right\}^{-1/2} - Z_{1-\alpha} \Bigg] . \quad (3)$$

Power formula

Approximation 2

Assuming $\frac{p_D^{j1}}{p_D^{j0}} \approx 1$ and $\frac{m_{*j0}}{n_{j0}} \approx \left[p_j(1-p_D^{j1}) + (1-p_j)(1-p_D^{j0})\right]$, the power of the test statistic is given by

$$\Phi\left[\sqrt{n}\,(\beta_1-\beta_0)\left\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_j p_D^{j0}+(1-p_j)p_D^{j1}\right]}\times\left(\left((1-p_j)+e^{\beta_j}p_j\right)^2 + \frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0}+p_j p_D^{j1}\right)^2}{\psi_j\left(e^{2\beta_j}p_j p_D^{j0}+p_D^{j1}\right)\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}\right)\right\}^{-1/2} - Z_{1-\alpha}\right]. \quad (4)$$

Power formula

Approximation 3

Assuming that the censoring variable is degenerate at τ with probability $1 - p_C$ and the approximation of the risk sets as $\frac{m_{*j0}}{n_{j0}} \approx (1 - p_C)\left[p_j(1 - p_D^{j1}) + (1 - p_j)(1 - p_D^{j0})\right]$, the power of the test statistic is given by

$$\Phi\left[\sqrt{n}\,(\beta_1 - \beta_0)\left\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_j p_D^{j0} + (1-p_j)p_D^{j1}\right]} \times \left(\left((1-p_j) + e^{\beta_j}p_j\right)^2 + \frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0} + p_j p_D^{j1}\right)^2}{(1-p_C)\psi_j\left(e^{2\beta_j}p_j p_D^{j0} + p_D^{j1}\right)\left(p_j(1-p_D^{j1}) + (1-p_j)(1-p_D^{j0})\right)}\right)\right\}^{-1/2} - Z_{1-\alpha}\right]. \quad (5)$$

Sample size formula

Theorem

For a given power ϑ, significance level α, and the denominator of any of the power formula ((3), (4) and (5)), denoted as σ_{den}, to detect the ratio of the hazard ratio, $exp(\beta_1 - \beta_0)$, for the treatment effect between the two biomarker groups, the required total cohort size is

$$\frac{(Z_\vartheta + Z_{1-\alpha})^2 \sigma^2_{den}}{(\beta_1 - \beta_0)^2}. \tag{6}$$

Simulation Results

Simulation

- Treatment is randomly assigned with probability 0.5 and biomarker proportions considered: 0.3, 0.5.
- Censoring time $\sim$ mixture distribution, with probability p_C from uniform distribution in $[0, \tau]$ and probability, $(1 - p_C)$ being degenerate at τ.
- Table 1 show that the Type I error of the test.
- Data generated from Weibull(2).
- Treatment group '0': $\lambda_{j0}(t) = 2\lambda_j t, \quad t \in (0, \infty)$.
- Treatment group '1': $\lambda_{j1}(t) = 2\lambda_j t e^{\beta_j}$, $t \in (0, \infty)$, $j = 0, 1$.
- $\lambda_0 = 1$ and $\lambda_1 = 0.75$, 1 and 1.25.
- Number of simulations = 20000.

Simulation

Table 1: Summary of Type I Error for Weibull (2) for $\beta_1 - \beta_0 = 0.25$ and $1 - p_C = 0.8$

Distribution	Event prop.	Biomarker prop.	Full Cohort	Case-Cohort	Sub-cohort
(1, 0.75)	(0.05)	0.3	0.0517	0.0518	0.0422
		0.5	0.0484	0.0501	0.0468
	(0.1)	0.3	0.0492	0.0460	0.0501
		0.5	0.0521	0.0494	0.0504
	(0.2)	0.3	0.0505	0.0513	0.0507
		0.5	0.0508	0.0495	0.049
(1, 1)	(0.05)	0.3	0.0502	0.0501	0.0419
		0.5	0.0493	0.0487	0.0406
	(0.1)	0.3	0.0498	0.0477	0.0485
		0.5	0.0486	0.0505	0.0484
	(0.2)	0.3	0.0482	0.0486	0.0494
		0.5	0.0485	0.0509	0.0474
(1, 1.25)	(0.05)	0.3	0.0489	0.0473	0.036
		0.5	0.0479	0.0490	0.0231
	(0.1)	0.3	0.0493	0.0512	0.0502
		0.5	0.0509	0.0509	0.0503
	(0.2)	0.3	0.0486	0.0499	0.0508
		0.5	0.0492	0.0503	0.0493

Simulation

- Data is generated from Weibull(k), k = 1, 2, 3.
- $p_C = 0.3, 0.2, 0.1$.
- The sample size considered is 4000; # of simulations: 5000.
- Treatment group '0': $\lambda_{j0}(t) = l\lambda_j t^{l-1}, \quad t \in (0, \infty)$.
- Treatment group '1': $\lambda_{j1}(t) = l\lambda_j t^{l-1} e^{\beta_j}$, $t \in (0, \infty)$, $j = 0, 1, l = 2, 3$.
- $\beta_0 = 0.5$, $\beta_1 = 1$ and $P(\text{Treatment} = 1 \mid \text{Biomarker} = j) = 0.5$.
- Compared the theoretical power with the empirical power.

Recommended formula

- Based on different theoretical power formulae, we have summarized:

Table 2: Summary of Recommended Power formula based on Simulations

p_C	p_D^{00}	Formula
(0, 0.1]	(0, 0.2)	(3)
	[0.2,1)	(4)
(0.1, 0.2]	(0, 0.1)	(5)
	[0.1, 0.2)	(3)
	[0.2,1)	(4)
(0.2, 1)	(0, 0.2)	(5)
	[0.2, 1)	(3)

Simulation

Table 3: Summary of Power Calculation for Weibull(3) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.7$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort		Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Empirical
(1, 0.75)	(0.05)	0.3	0.549	0.406	0.397	0.145
		0.5	0.620	0.465	0.442	0.138
	(0.1)	0.3	0.780	0.527	0.482	0.217
		0.5	0.842	0.589	0.537	0.215
	(0.2)	0.3	0.964	0.589	0.6	0.328
		0.5	0.982	0.667	0.659	0.364
(1, 1)	(0.05)	0.3	0.587	0.434	0.413	0.159
		0.5	0.664	0.488	0.465	0.173
	(0.1)	0.3	0.808	0.516	0.494	0.217
		0.5	0.882	0.592	0.554	0.249
	(0.2)	0.3	0.97	0.603	0.599	0.348
		0.5	0.989	0.675	0.657	0.391

Simulation

Table 4: Summary of Power Calculation for Weibull(3) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.9$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort		Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Empirical
(1, 0.75)	(0.05)	0.3	0.561	0.436	0.43	0.157
		0.5	0.622	0.478	0.477	0.148
	(0.1)	0.3	0.808	0.559	0.557	0.219
		0.5	0.862	0.624	0.582	0.233
	(0.2)	0.3	0.972	0.653	0.63	0.333
		0.5	0.988	0.698	0.697	0.375
(1, 1)	(0.05)	0.3	0.582	0.447	0.447	0.145
		0.5	0.657	0.502	0.504	0.161
	(0.1)	0.3	0.832	0.562	0.561	0.224
		0.5	0.899	0.631	0.64	0.256
	(0.2)	0.3	0.980	0.652	0.634	0.366
		0.5	0.991	0.715	0.705	0.407

Simulation

- Compare empirical power based on the calculated sample size with the expected theoretical power.
- Power: 80%.
- $p_C = 0.15$.
- $\beta_1 = 1$ and $\beta_0 = 0.5$.
- $P(\text{Treatment} = 1 \mid \text{Biomarker} = j) = 0.5$.

Simulation

Table 5: Summary of Sample Size and Empirical Power for Theoretical Power = 80% using Table 2

Event prop.(p_D^{00})	Biomarker prop.	Subcohort prop.	Exponential n	Exponential Empirical Power	Weibull(2) n	Weibull(2) Empirical Power
0.05	0.3	0.1	12617	0.8	12616	0.8192
		0.2	9657	0.8088	9657	0.8048
	0.4	0.1	10949	0.8244	10947	0.808
		0.2	8356	0.804	8356	0.808
	0.5	0.1	10423	0.8224	10421	0.816
		0.2	7932	0.814	7932	0.7988
0.1	0.3	0.1	8797	0.818	8791	0.8032
		0.2	5951	0.816	5950	0.8064
	0.4	0.1	7703	0.8244	7696	0.8148
		0.2	5184	0.7916	5182	0.8016
	0.5	0.1	7399	0.814	7392	0.8192
		0.2	4953	0.8144	4951	0.7976

Simulation

Table 6: Summary of Sample Size and Empirical Power for Theoretical Power = 80% using (5)

Event prop.(p_D^{00})	Biomarker prop.	Subcohort prop.	Exponential n	Exponential Empirical Power	Weibull(2) n	Weibull(2) Empirical Power
0.05	0.3	0.1	12617	0.8	12616	0.8192
		0.2	9657	0.8088	9657	0.8048
	0.4	0.1	10949	0.8244	10947	0.808
		0.2	8356	0.804	8356	0.808
	0.5	0.1	10423	0.8224	10421	0.816
		0.2	7932	0.814	7932	0.7988
0.1	0.3	0.1	9385	0.8344	9378	0.82
		0.2	6212	0.8292	6210	0.8192
	0.4	0.1	8175	0.8376	8167	0.832
		0.2	5393	0.8188	5391	0.8244
	0.5	0.1	7812	0.8416	7803	0.8396
		0.2	5137	0.8168	5134	0.8268

Cost Efficiency of Case-Cohort Design

- Cost-efficiency of case-cohort compared to SRS.
- It is the ratio of the sample sizes required in the two sampling schemes to attain the same power ϑ.
- SRS sample size $= n^*_{SRS} = \frac{(Z_\vartheta + Z_{1-\alpha})^2 \times \sigma^2_{\text{den}_{SRS}}}{(\beta_1 - \beta_0)^2}$.
- (6) is the samplesize formula for CC design.

Cost Efficiency Ratio

The ratio of the two is given by $R = \frac{\sigma^2_{\text{den}_{SRS}}}{\psi\sigma^2_{\text{den}} \sum_{j=0}^{1} \left[r_j \left\{ 1 + (\frac{1-\psi}{\psi}) \left(p_j p_D^{j1} + (1-p_j) p_D^{j0} \right) \right\} \right]}$ for fixed total cohort size n and assuming that the sub-cohort proportion , ψ_j is the same in the two biomarker groups.

Cost Efficiency of Case-Cohort Design

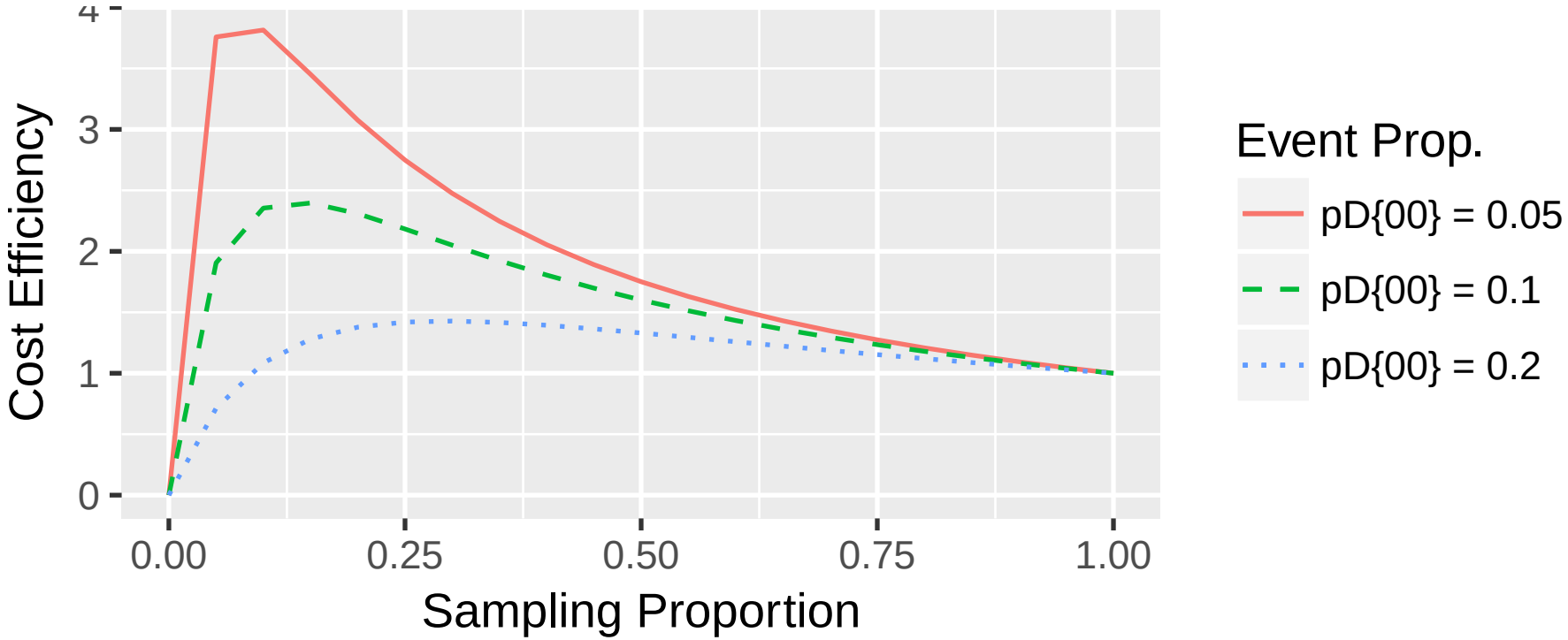


Figure 1: The Cost-Efficiency Curve of the Case-Cohort Design

Application to LACE & CALGB data

- Two-phase design based on combined LACE & CALGB 9633 databases (Shepherd et al., 2014).
- Eligible patients = 1422.
- Patients with KRAS-wild-type biomarker(1) = 1146 (80.6%)
  - Patients in ACT arm(1) = 581 (50.7%)
  - Patients in OBS arm(0) = 565 (49.3%)
- Patients with KRAS-mutated biomarker(0) = 276 (19.4%)
  - Patients in ACT arm(1) = 143 (51.8%)
  - Patients in OBS arm(0) = 133 (48.2%)
- $p_D^{00} = 0.044$, $p_D^{01} = 0.065$, $p_D^{10} = 0.098$, $p_D^{11} = 0.035$.
- $exp(\hat{\beta}_1) = 0.32$ and hazard ratio = 0.25.
- For power 60%, 70% and 75%, the subcohort sample sizes are 143, 347 and 837.

Discussion

Discussion and Future Work

- A log-rank type test statistic is considered.
- Explicit formulas are obtained for the calculation of power and sample size.
- Finite sample results show for low disease incidence, the design produces fairly high power.
- Proposed formula is more cost efficient than the simple random sample.
- Binary biomarker is considered.

- Consider Discrete (levels > 2) or continuous biomarker groups.
  - Develop test statistics to test effects being equal in all groups.
- Stratified two-phase design

Thank you!

Thank you!