TWO-PHASE DESIGNS FOR TIME-TO-EVENT DATA

National University of Singapore
July 4th 2022

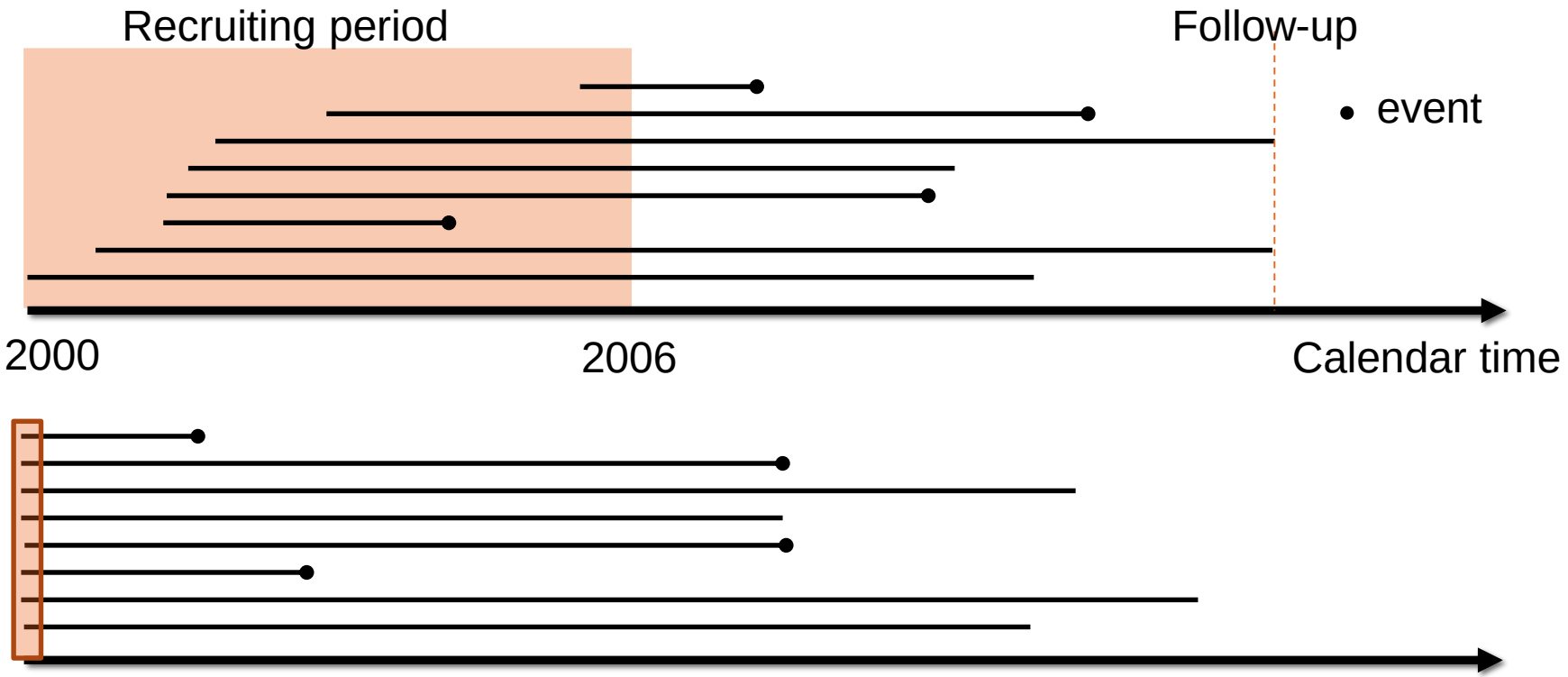
Paola Rebora
Bicocca Bioinformatics, Biostatistics and Bioimaging Centre - B4,
Department of Medicine and Surgery,
University of Milano-Bicocca, Monza, Italy

Two-stage studies and time-to-event data:

- Two-stage designs could be particularly useful in cohort studies with time-to-event end-points.
- For example to identify new biomarkers.
- In fact cohort studies often have stored biologic samples and follow-up over many years and will require **efficient study designs** for parsimonious use of specimens and to limit costs of biological analyses.

Example :

Clinical trial (AIEOP ALL-2000) on 1999 children with acute lymphoblastic leukemia (ALL). Diagnosed from 2000 to 2006.
Bio-bank with samples at diagnosis.



AIEOP-BFM ALL 2000 study - Conter, et al. Blood 2010 115:3206-3214;

Example :

Clinical trial (AIEOP ALL-2000) on 1999 children with acute lymphoblastic leukemia (ALL). Diagnosed from 2000 to 2006.
Bio-bank with samples at diagnosis.

Cytosolic glutatione S-transferasi (GST) genes involved in drug metabolism. DELETION should increase availability of anticancer drugs
GST-T1 (deletion in 13%-26% of Caucasian population)
Unknown regulatory role

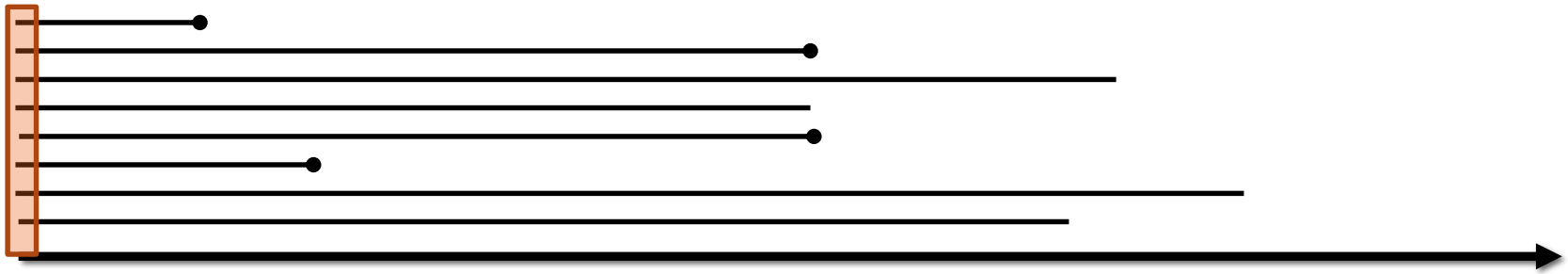
AIM: to investigate the influence of GST-T1 on treatment failure due to relapse.

Franca R, Rebora P, Basso G et al. Pharmacogenomics 2012;13:1905-16.

Example of two-phase study :

Clinical trial (AIEOP ALL-2000) on 1999 children with acute lymphoblastic leukemia (ALL). Diagnosed from 2000 to 2006.
Bio-bank with samples at diagnosis.

Clinical trial cohort (N=1999) with clinical informations and outcome



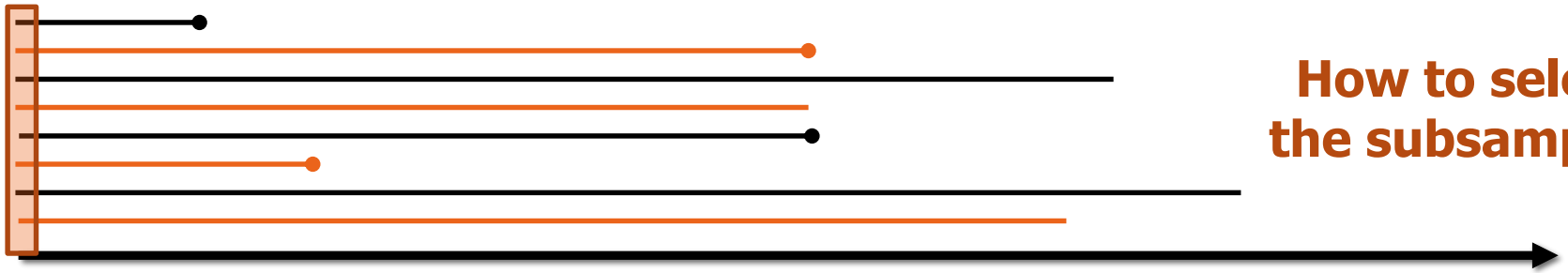
Biological samples stored at diagnosis

Example of two-phase study :

Clinical trial (AIEOP ALL-2000) on 1999 children with acute lymphoblastic leukemia (ALL). Diagnosed from 2000 to 2006.
Bio-bank with samples at diagnosis.

Clinical trial cohort (N=1999) with clinical informations and outcome

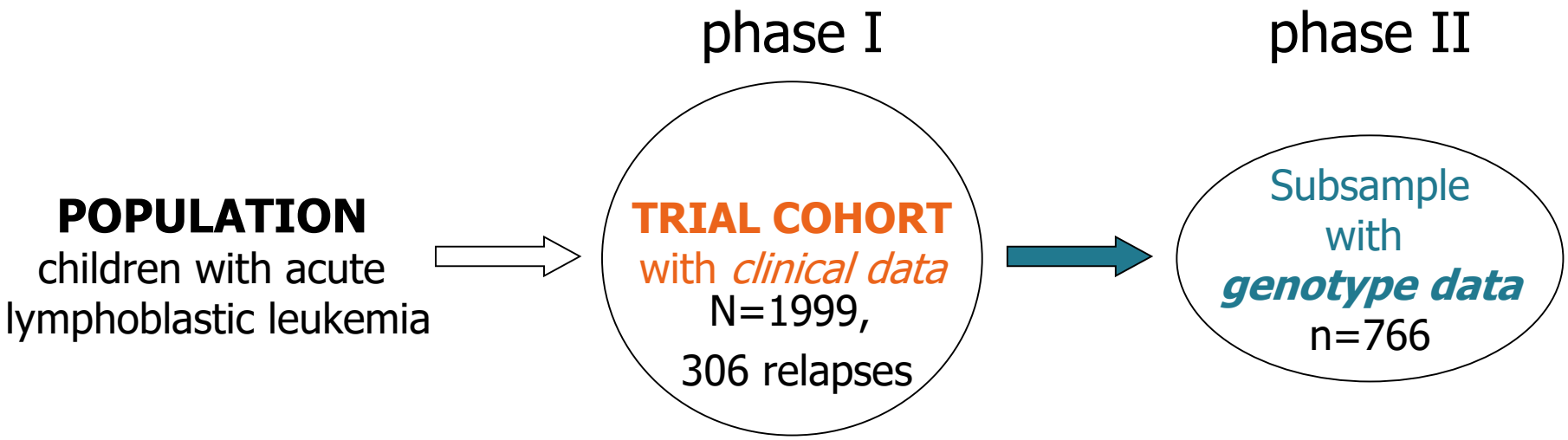
Subsample (n) on which to measure the biomarker



How to select the subsample?

Biological samples stored at diagnosis

Example of two-phase study :



How to select an «informative» subsample on which to measure the biomarker (i.e. GST deletion)?

OPTIMAL SAMPLING OF 2-STAGE DATA (lecture 4.2)

Example of stratified two-phase study :

Clinical trial on 1999 children with ALL (phase I)

Full cohort	Treatment/risk stratification			
	standard	medium	high	TOT
No relapse	487	987	219	**1693**
Relapse	28(5%)	186(16%)	92(30%)	**306**
TOT	515	1173	311	**1999**

Subsample	Treatment/risk stratification			
	standard	medium	high	TOT
No relapse	?	?	?	
Relapse	?	?	?	
TOT				**766**

Minimum potential follow-up 2 years

Which sampling fraction?

Optimal Sampling Strategy for two-stage studies

(Reilly, AJE 1996)

To get the highest efficiency and thus MINIMIZE the variance of the coefficient of interest (marker)

⇓

the sampling fraction for each stratum should be proportional to the variability within the stratum as compared with total variability

⇓

sample more data from strata with higher variability

→ need pilot data

Pilot data:

AVAILABLE DATA	Risk/trt stratification			
	standard	medium	high	
Not relapse	22	53	8	83
Relapses	14	70	1 +1	85
Tot	36	123	9	168

*This strata was very poor represented in the available data (in order to include it I artificially introduced a faked observation in this strata with a different value for the SNP, so that it showed variability within the strata).

Optimal sampling as if binary data

We applied the function **optfixn** to select an optimal second-stage sample of a fixed size (n=766):

OPTIMAL SAMPLING FRACTIONS	Treatment/risk stratification			
	standard	medium	high	
No relapse	65(0.13)	255(0.26)	140(0.64)	460
Relapse	28 (1)	186 (1)	92 (1)	306
	93	441	232	766

All relapses were sampled (typically).

460 "no relapses" were randomly drawn from the cohort according to the optimal sampling fractions.

Subsample – genotyped data:

ACTUAL SAMPLING FRACTIONS	Treatment/risk stratification			
	standard	medium	high	
No relapse	54(0.11)	193(0.20)	109(0.50)	356
Relapse	21(0.75)	147(0.79)	77(0.84)	245
	75	340	186	601

The lab genotyped biological samples of 601 patients:

- 107 with deleted GST-T1, 48 relapses
- 494 not deleted GST-T1, 197 relapses

```
Meanscore estimate
```

OR (GST-T NULL vs NORM)=1.19 (95%CI: 0.73; 1.83)

Relapse incidence by GST-T: weighted versus unweighted estimates

GST-T type	# at risk	# relapses	Relapse incidence	Weighted relapse incidence
NULL	107	48	48/107=44.9%	18.3%
NORM	494	197	197/494=39.9%	14.7%

Efficient design requires appropriate analysis!

Relapse incidence by GST-T: weighted in GST-T deleted subjects (NULL)

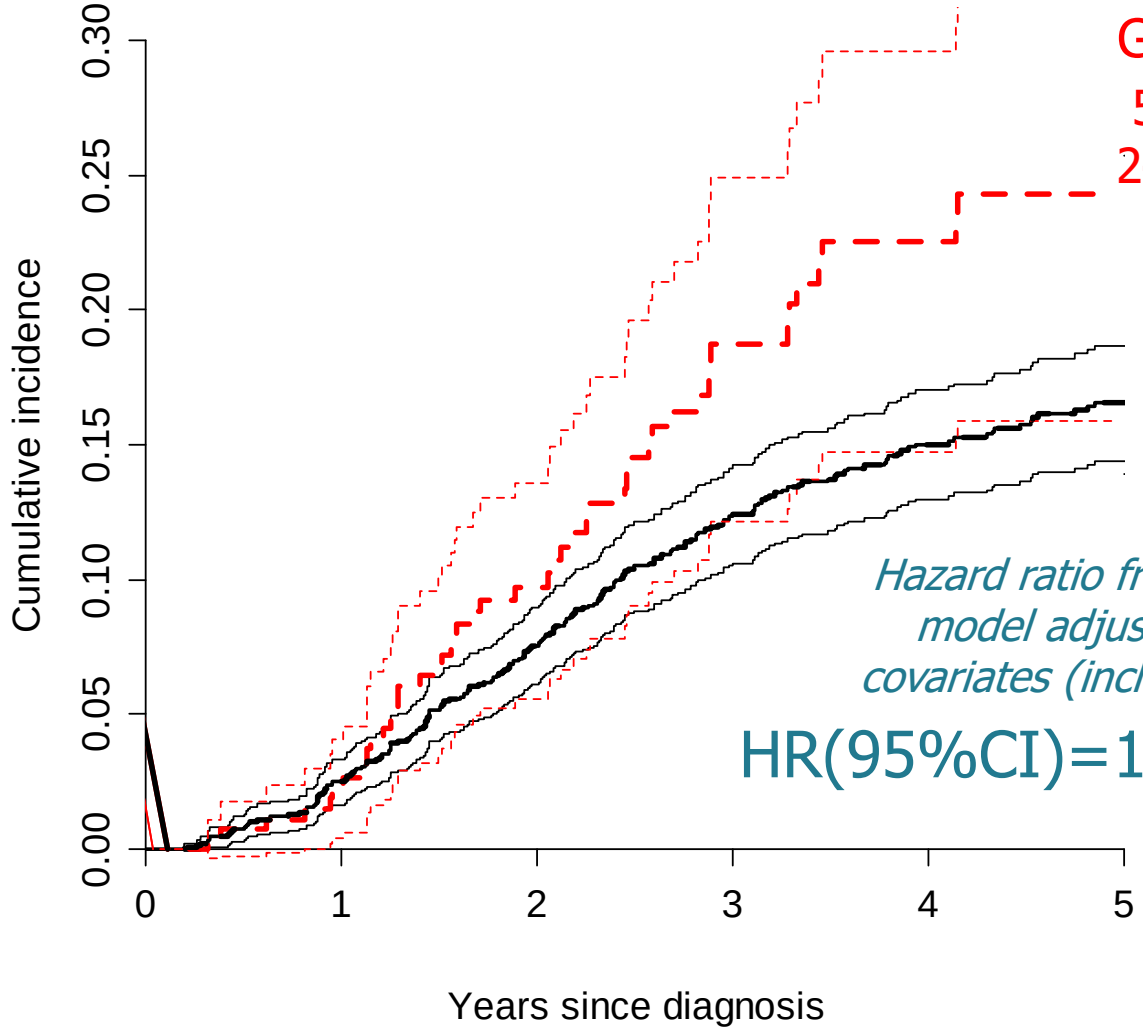
GST-T NULL subsample

SAMPLING FRACTIONS	Treatment/risk stratification			
	standard	medium	high	
No relapse	6(0.11)	34(0.20)	19(0.50)	59
Relapse	**5(0.75)**	**25(0.79)**	**18(0.84)**	**48**
	11	59	37	107

Weighted relapse incidence in GST-T deleted subjects (NULL):

$$\frac{5 * \frac{1}{0.75} + 25 * \frac{1}{0.79} + 18 * \frac{1}{0.84}}{5 * \frac{1}{0.75} + 25 * \frac{1}{0.79} + 18 * \frac{1}{0.84} + 6 * \frac{1}{0.11} + 34 * \frac{1}{0.2} + 19 * \frac{1}{0.5}} = \frac{59}{322} = 18.3\%$$

Relapse incidence* by GST-T:



GST-T NULL

5-yrs incidence:
24.4%(16.1-31.9)

GST-T NORM

5-yrs incidence:
16.5%(14.4-18.6)

Hazard ratio from weighted Cox model adjusted by relevant covariates (including risk groups)

HR(95%CI)=1.34(0.90-2.00)

* Rebora P, Valsecchi MG. Survival estimation in two-phase cohort studies with application to biomarkers evaluation. *Stat Methods Med Res.* 2014 May 19

Weighted Cox model for two-phase studies:

- Weighed partial likelihood where weights are reciprocal of sampling fractions
- variance (adjusted for sampling) can be split in two terms denoting variation due to:
  1. sampling of phase I (0.022) - Estimate of the minimum irreducible uncertainty for the cohort
  2. sampling of phase II from phase I (0.019) - Remain due to genotyping only the subsample

$$efficiency = \frac{0.022}{0.022 + 0.019} = 0.54$$

54% of efficiency genotyping 30% (601/1999) of the sample!

Lin D-Y (2000) Biometrika 87: 37-47

Survey package in R by Thomas Lumley

Optimal sampling with time-to-event endpoint

- Optimal sampling applied *as if* binary data
- We got 54% of efficiency genotyping 30% (601/1999) of the sample!
- Data analyses with survival methods (weighted)

Design of stratified two-phase studies

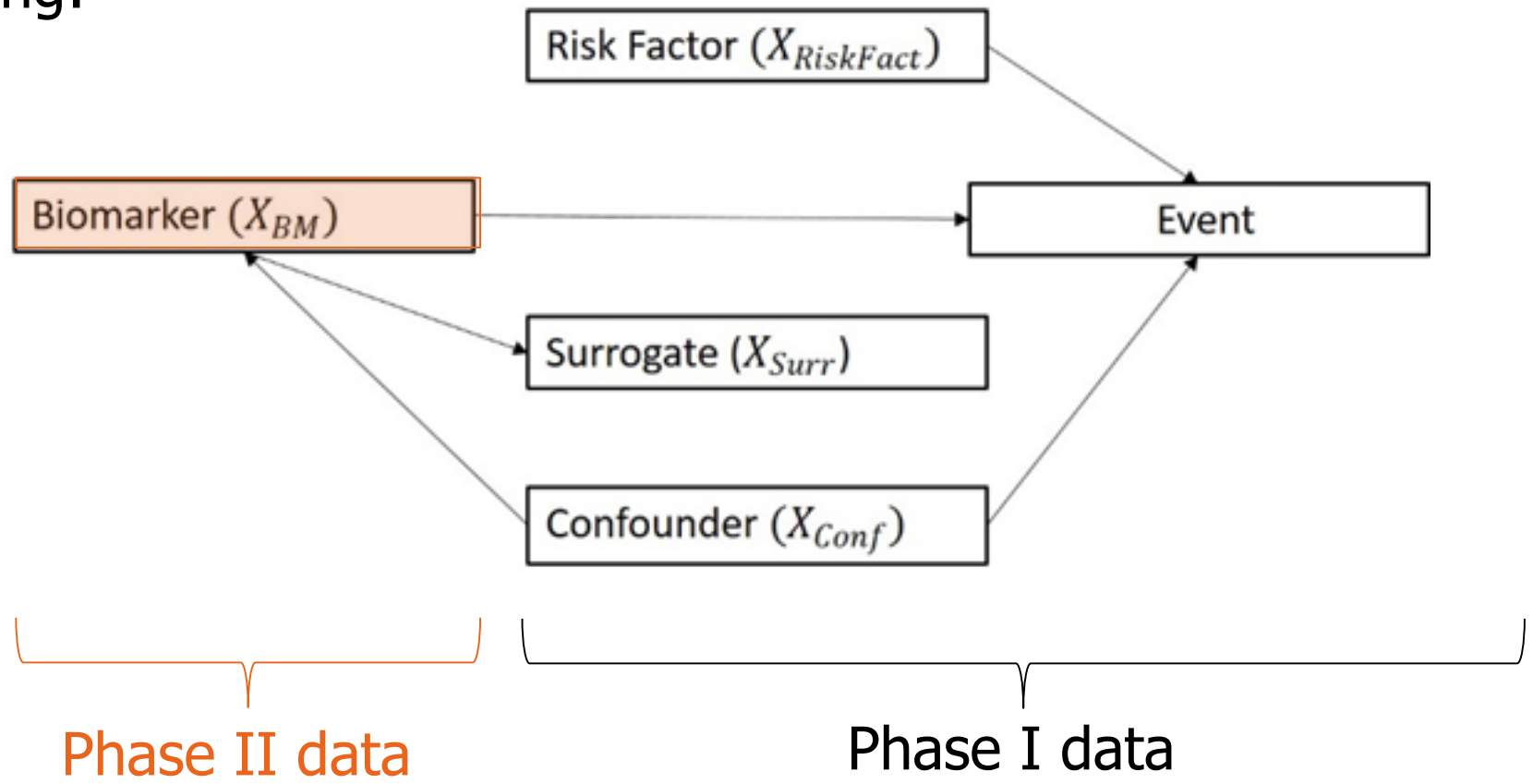
- Size of the subsample is often driven by budget constrains, but it is important to assess power
- Which variables should we use for stratification?
- What if pilot data are not available?

Simulations can help in comparing the performance of different designs and to estimate power in complex settings

Graziano, Valsecchi & Rebora BMC-MRM 2021 Sampling strategies for a prognostic biomarker

Design of stratified two-phase studies

Setting:



Design of stratified two-phase studies

Simulation (2000 simulations of phase I data with N=2000)

We mimic different sampling scheme for second phase data with a fixed size n (Biomarker measurment):

- Simple Random Sample (SRS)
- Probability Proportional to Size (PPS)
- Case-Control (CC)
- Stratified CC

With strata defined using the following variables:
- event,
- event and risk factor,
- event and confounder,
- event and surrogate.

Weighted Cox model used to assess the influence of the biomarker on the event (adjusting for the confounder)

Results of simulations

Table 1 Bias, empirical standard error, mean square error, power and design effect of the biomarker regression coefficient estimate ($\hat{\beta}_{BM}$) for the full cohort and different sampling designs. Accuracy of surrogate: sensitivity (i.e. probability of having a positive surrogate if the biomarker is positive) = 0.7 and specificity (i.e. probability of having a negative surrogate if the biomarker is negative) = 0.7, biomarker common (a) and rare (b)

Sampling design	Stratification variable	n*	Bias			SE empirical			MSE			Power (%)			Design effect		
			Censoring rate			Censoring rate			Censoring rate			Censoring rate			Censoring rate		
			0	0.1	0.4	0	0.1	0.4	0	0.1	0.4	0	0.1	0.4	0	0.1	0.4
a)																	
Full cohort	–	2000	0.008	−0.015	0.009	0.093	0.095	0.112	0.009	0.009	0.013	99	97	95	–	–	–
1. SRS	–	600	0.004	−0.013	0.006	0.182	0.187	0.206	0.033	0.035	0.042	64	58	53	–	–	–
2. PPS	Event	599	0.007	−0.015	0.007	0.173	0.180	0.199	0.029	0.033	0.039	65	58	54	1.003	1.003	1.005
2a. PPS	Event; Risk factor	598	0.008	−0.016	0.004	0.172	0.175	0.205	0.029	0.031	0.042	65	58	52	1.002	1.003	1.002
2b. PPS	Event; Confounder	598	0.003	−0.015	0.002	0.174	0.179	0.203	0.030	0.032	0.041	65	57	51	0.999	1.002	1.000
2c. PPS	Event; Surrogate	598	0.007	−0.013	0.013	0.161	0.171	0.190	0.026	0.029	0.036	69	64	57	1.106	1.129	1.104
3. CC	Event	600	0.011	−0.008	0.019	0.159	0.158	0.179	0.025	0.025	0.032	74	68	67	1.179	1.219	1.352
3a. CC	Event; Risk factor	600	0.010	−0.009	0.008	0.162	0.166	0.182	0.026	0.028	0.033	72	65	62	1.139	1.176	1.307
3b. CC	Event; Confounder	600	0.012	−0.015	0.010	0.162	0.161	0.175	0.026	0.026	0.031	73	65	66	1.182	1.187	1.354
3c. CC	Event; Surrogate	600	0.008	−0.016	0.012	0.148	0.153	0.170	0.022	0.024	0.029	76	71	69	1.334	1.363	1.495
4. NCC	Event	550	0.008	−0.018	0.014	0.169	0.165	0.175	0.029	0.028	0.031	68	63	67	1.066	1.144	1.378
5. CM	Event; Surrogate	546	−0.044	−0.058	−0.009	0.151	0.153	0.165	0.025	0.027	0.027	67	61	67	1.379	1.395	1.536

B=2000 simulations SURROGATE/AUXILIARY = Sens 0.7 and spec 0.7, n=600 and BM frequency= 25% β_{BM} = 0.28 (HR=1.323)

Results of simulations: surrogate/auxiliary

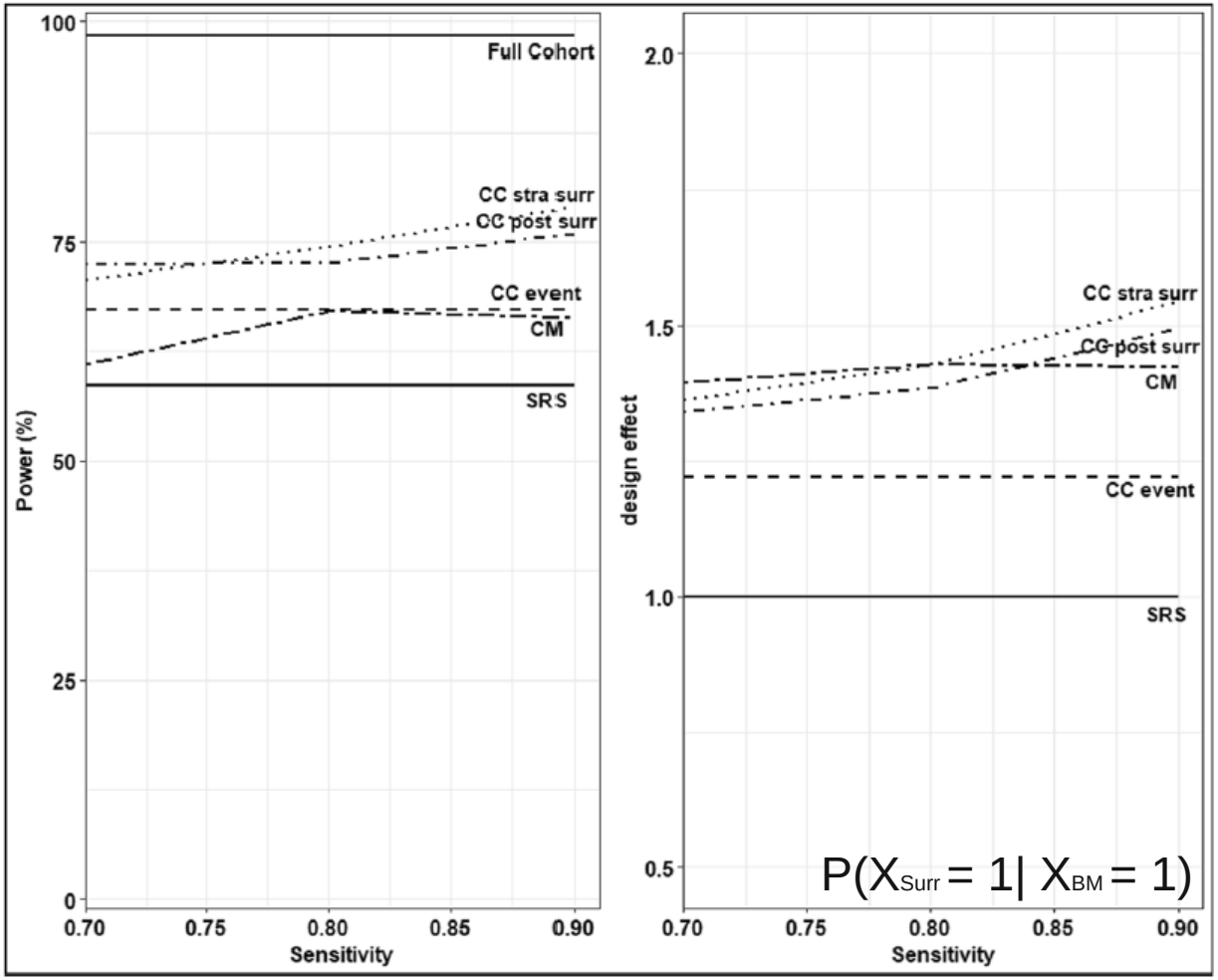


Fig. 3 Power and design effect for different sensitivity levels (i.e. probability of having a positive surrogate if the biomarker is positive) of the surrogate variable. Scenario: specificity (i.e. probability of having a negative surrogate if the biomarker is negative) =0.7, censoring rate $\rho = 0.1$, hazard ratio of biomarker =1.5 and sample size of phase II (n) =600. Legend: CC stra surr (Case-Control stratified by surrogate), CC post surr (Case-Control post stratified by surrogate), CC event (Case-Control), CM (Counter-Matching) and SRS (simple random sampling)

Leukemia data: efficiency comparison

In the example of the 2-stage design on ALL clinical trial we estimated a 54% of efficiency genotyping 30% (n=601 / N=1999) of the sample by the optimal sampling strategy.

By simulation we can estimate the efficiency of different designs (with respect to the full cohort, n=N):

Table 2 Efficiency (refers to the full cohort), design effect (refers to Simple Random Sampling) and power for SRS and Case-Control (CC) designs with hypothetical hazard ratio of the biomarker of interest (HR_{BM}) of 1.3 and 1.5, biomarker common (25%), censoring rate $\rho = 0.1$, type I error 0.05

	SRS	Case-control	CC stratified by surrogate	CC stratified by risk factor
Efficiency				
$HR_{BM} = 1.3$	30.40%	38.91%	43.06%	34.47%
$HR_{BM} = 1.5$	25.98%	36.26%	38.51%	32.73%
Design effect				
$HR_{BM} = 1.3$	–	1.23	1.37	1.20
$HR_{BM} = 1.5$	–	1.22	1.36	1.18
Power				
$HR_{BM} = 1.3$	30.91%	54.80%	60.15%	54.34%
$HR_{BM} = 1.5$	58.35%	68.10%	70.65%	65.40%

$$\text{efficency} = \frac{\text{variance } \textit{of the full cohort} \text{ (n=N=1999)}}{\text{variance in the subsample with size n=601}}$$

Leukemia data: power estimate

Table 2 Efficiency (refers to the full cohort), design effect (refers to Simple Random Sampling) and power for SRS and Case-Control (CC) designs with hypothetical hazard ratio of the biomarker of interest (HR_{BM}) of 1.3 and 1.5, biomarker common (25%), censoring rate $\rho = 0.1$, type I error 0.05

	SRS	Case-control	CC stratified by surrogate	CC stratified by risk factor
Efficiency				
$HR_{BM} = 1.3$	30.40%	38.91%	43.06%	34.47%
$HR_{BM} = 1.5$	25.98%	36.26%	38.51%	32.73%
Design effect				
$HR_{BM} = 1.3$	–	1.23	1.37	1.20
$HR_{BM} = 1.5$	–	1.22	1.36	1.18
Power				
$HR_{BM} = 1.3$	30.91%	54.80%	60.15%	54.34%
$HR_{BM} = 1.5$	58.35%	68.10%	70.65%	65.40%

Power can be estimated by simulation by the design2phase package implemented in R software

Summary

- Optimal design could be efficently applied also to time-to-event data (need to work on ad-hoc optimal design)
- Comparison of different sampling strategies could be done by simulations to evaluate pro/cons in the specific setting
- Power estimate can be achieved by simulations

Referencees:

•Franca et al. Pharmacogenomics 2012 A novel (efficient) study design for time-to-event data

•Graziano Valsecchi & Rebora Sampling strategies for a prognostic biomarker. BMC-MRM 2021

•Breslow NE, Lumley T, Ballantyne CM, et al. Using the Whole Cohort in the Analysis of Case-Cohort Data. *AJE* 2009; 169(11):1398 – 1405

•Lumley T. *Complex Surveys: A Guide to Analysis Using R*. Wiley Series in Survey Methodology, John Wiley & Sons, 2010.

•Reilly M. Optimal sampling strategies for two-stage studies. *AJE* 1996; 143(1):92–100.

•Rebora P, Valsecchi MG. Survival estimation in two-phase cohort studies with application to biomarkers evaluation. *Stat Methods Med Res.* 2014 May 19.

•Rebora P, Antolini L, Glidden DV, Valsecchi MG. Crude incidence in two-phase designs in the presence of competing risks. BMC Med Res Methodol. 2016 Jan 11;16:5. doi: 10.1186/s12874-015-0103-1.