Adaptive signature designs for cancer vaccines

Andrea Callegaro
-GSK Vaccines, Belgium

PSI Conference 02 – 05 Jun, 2019

Disclaimer

- I am an employee of, and hold shares in, the GSK group of companies

Outline

1. Background of MAGE-A3 predictive gene signature and Adaptive Signature Design in Phase III studies

2. Statistical aspects to build the gene-signature
    1. Survival Models
    2. High-dimensional data
    3. Parametrization of the gene-treatment interaction
    4. method to find the cut-off

3. Gene-signature results in DERMA Phase III training set

4. Adaptive Signature Design with futility

5. Conclusions

MAGE-A3 cancer immunotherapy
Introduction and background

Antigen-Specific Cancer Immunotherapy

Antigen
- Recombinant proteins
- Combined with GSK's proprietary immunostimulants (Adjuvant Systems)
- Minimal implementation constraints
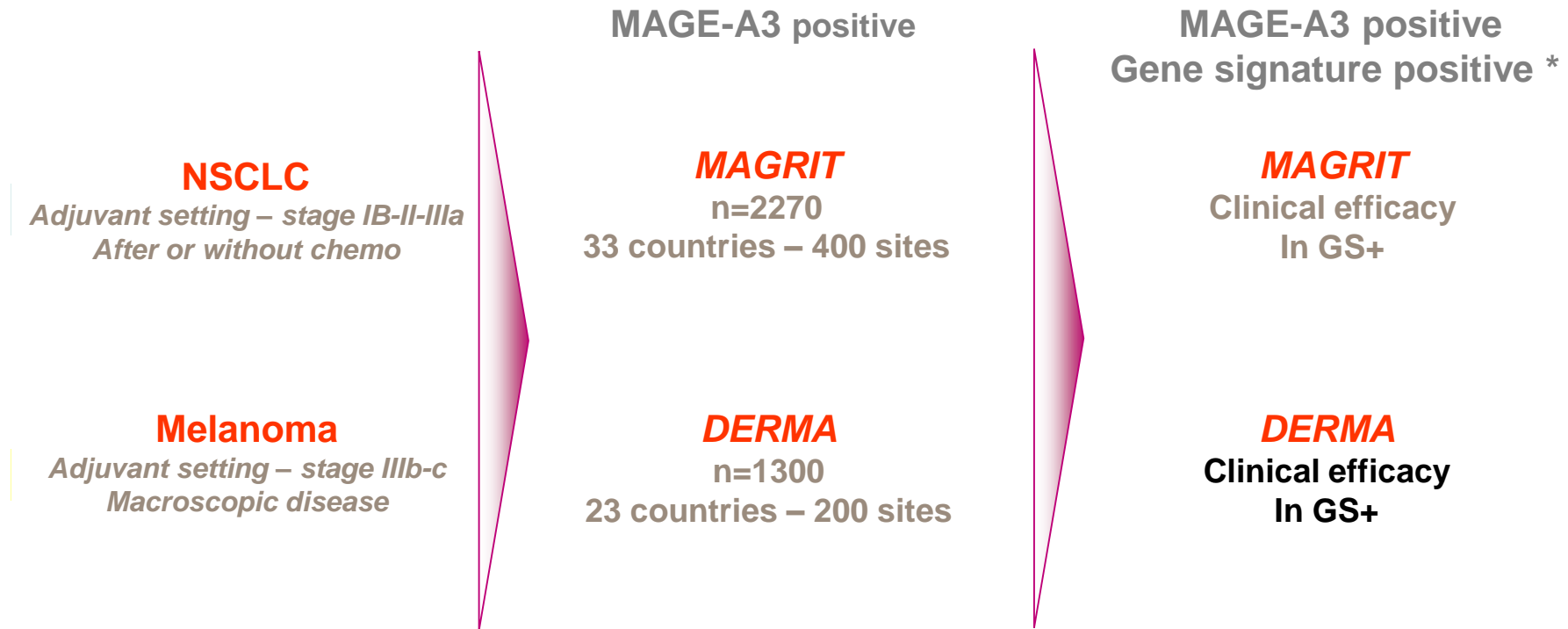
Specific
- Tumor cell specific*

Cancer
- Prevention of relapse
- Minimal residual disease

Immuno-therapy
- Educate the patient's immune system to fight cancer
- Novel approach involving all immune anti-cancer cells

Implementation of predictive biomarkers in Phase III clinical studies

gsk

MAGE-A3 positive

MAGE-A3 positive
Gene signature positive *

NSCLC
Adjuvant setting – stage IB-II-IIIa
After or without chemo

MAGRIT
n=2270
33 countries – 400 sites

MAGRIT
Clinical efficacy
In GS+

Melanoma
Adjuvant setting – stage IIIb-c
Macroscopic disease

DERMA
n=1300
23 countries – 200 sites

DERMA
Clinical efficacy
In GS+

* Gene signature predictive of clinical response and correlating with immune environment in the tumor

Biomarkers for Selection of Patients More Likely to Benefit from MAGE-A3 Immunotherapy:
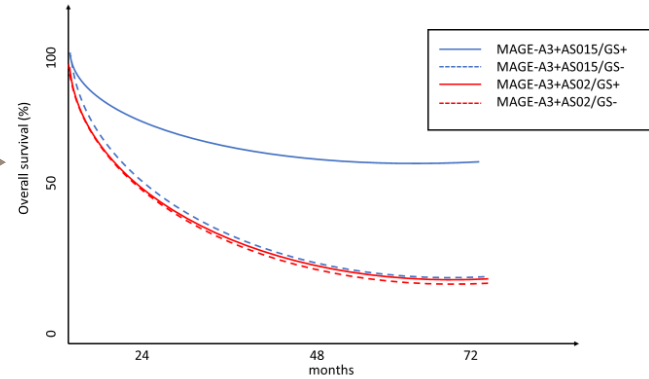# From Translational Research to Clinical Practice

Identification in Phase II unresect. metastatic melanoma

Frozen specimen/microarrays

100 probesets (84 genes)

HR (GS+ v GS–)
AS15: 0.37 (95% CI, 0.13-1.05; P =.06)
AS02B: 0.84 (95% CI, 0.36-1.97; P =.70)

Confirmation in Phase II adjuvant NSCLC

Frozen specimen/qRT-PCR
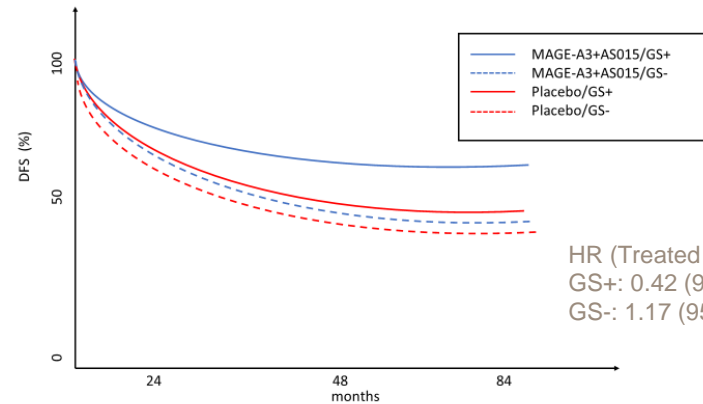
Melanoma classifier
61 genes
measured by qRTPCR

HR (Treated v placebo)
GS+: 0.42 (95% CI, 0.17-1.03; P =.06)
GS-: 1.17 (95% CI, 0.59-2.31; P=.65)

Ulloa-Montoya et al JCO, 2013

Biomarkers for Selection of Patients More Likely to Benefit from MAGE-A3 Immunotherapy:
From Translational Research to Clinical Practice

Identification in Phase II unresect. metastatic melanoma

Frozen specimen/microarrays

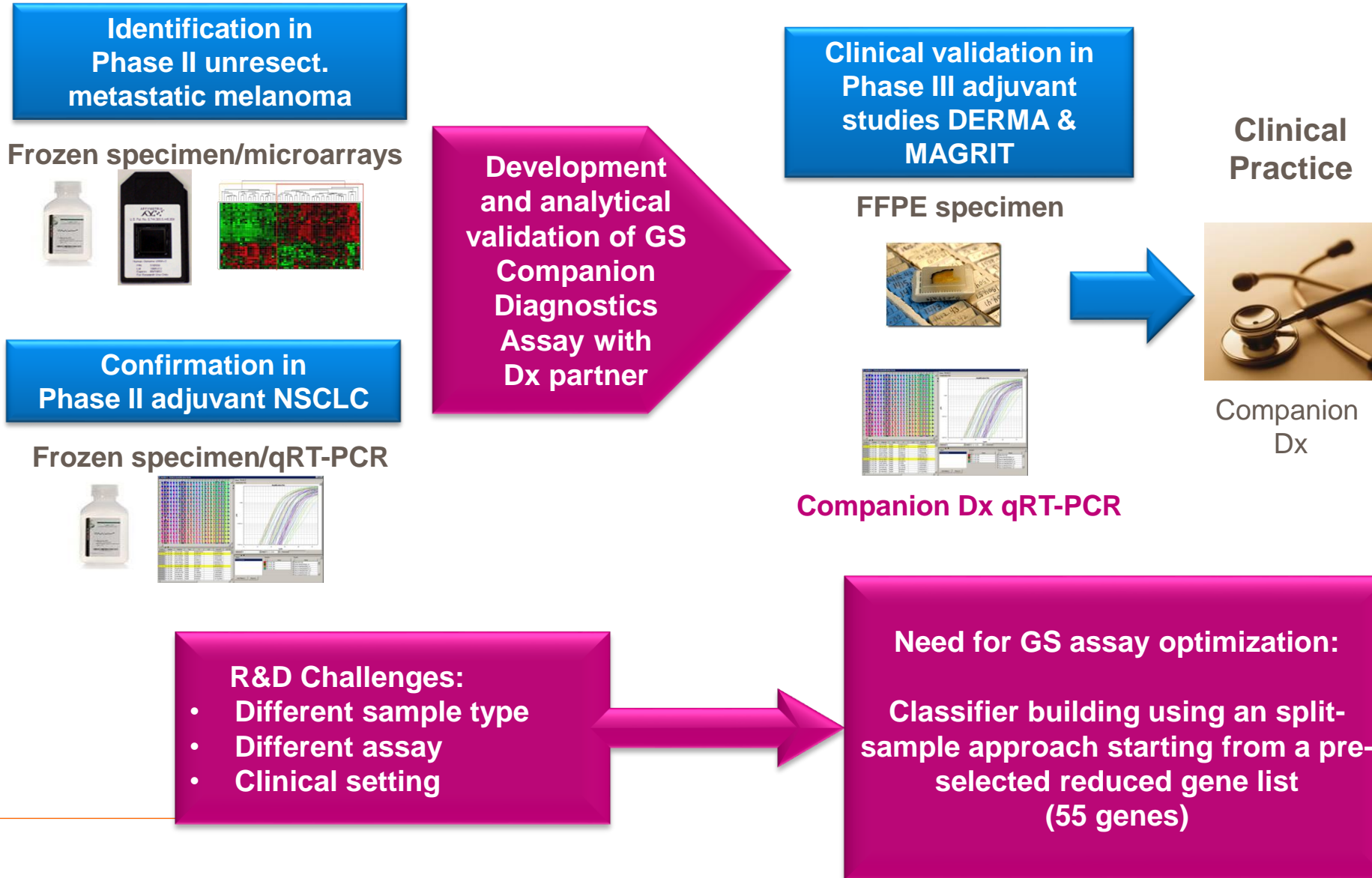**Confirmation in Phase II adjuvant NSCLC**

Frozen specimen/qRT-PCR

Development and analytical validation of GS Companion Diagnostics Assay with Dx partner

Clinical validation in Phase III adjuvant studies DERMA & MAGRIT

FFPE specimen

Companion Dx qRT-PCR

Clinical Practice

Companion Dx

R&D Challenges:
- **Different sample type**
- **Different assay**
- **Clinical setting**

Need for GS assay optimization:

Classifier building using an split-sample approach starting from a pre-selected reduced gene list (55 genes)
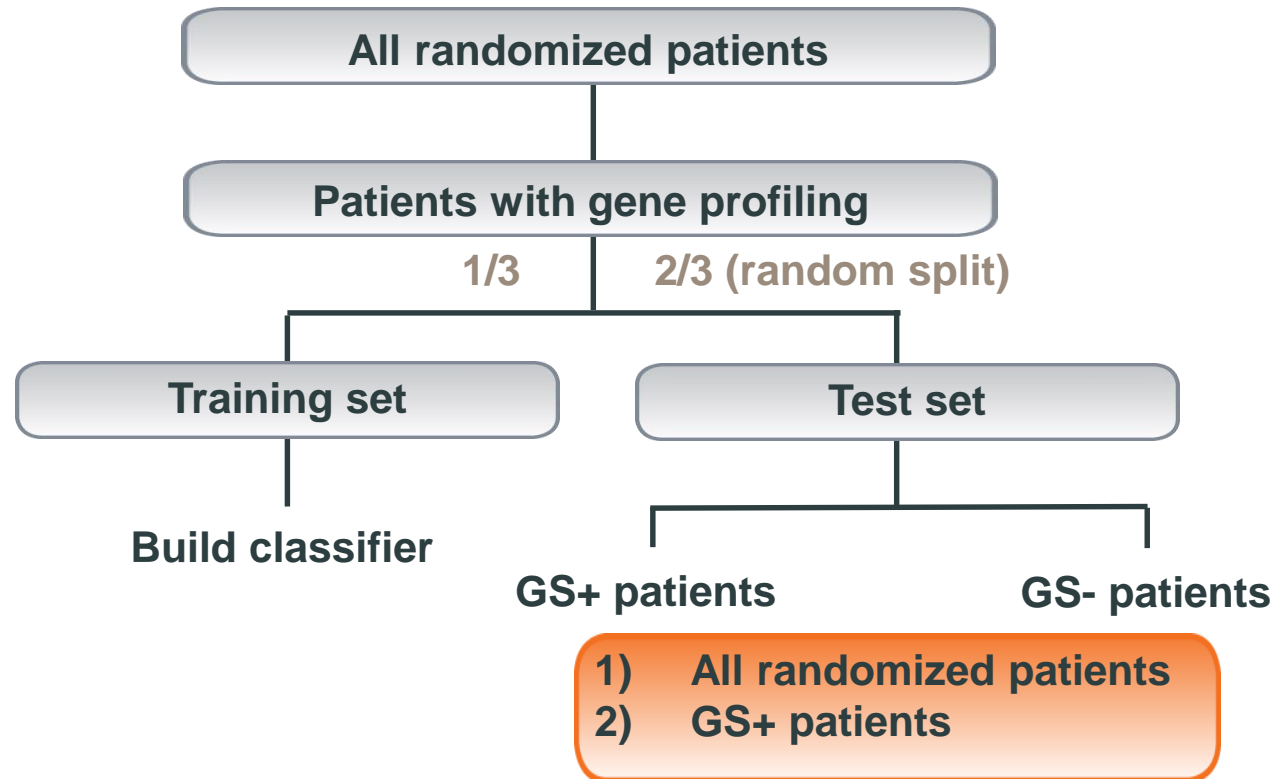
Adaptive Signature Design
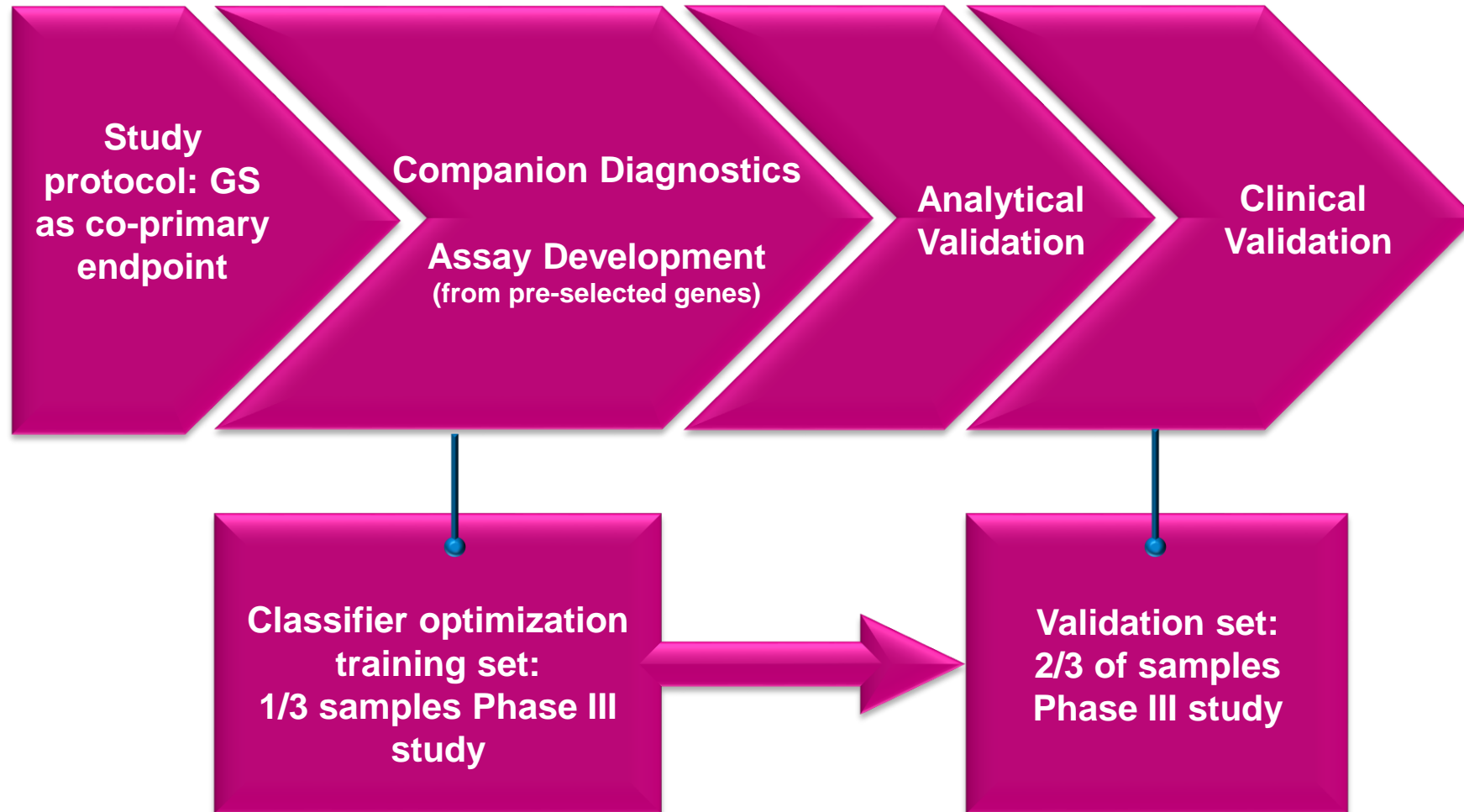## Prospective clinical validation of gene signature

- Freidlin and Simon: Adaptive Signature Design

- Change in sample type and methodology

Freidlin and Simon (2005) Clin Cancer Res; **11**, 7872-7878

Split-Sample Approach: Assay Optimization and Clinical Validation of the GS Biomarkers in the Ongoing Phase III MAGE-A3 Immunotherapy Studies

Study protocol: GS as co-primary endpoint

Companion Diagnostics

Assay Development (from pre-selected genes)

Analytical Validation

Clinical Validation

Classifier optimization training set: 1/3 samples Phase III study

Validation set: 2/3 of samples Phase III study

How to build the gene-signature

Supplementary appendix of
Dreno, B et al. (2018). Clin Cancer Res; 11: 7872–7878.
Li, J et al. (2016). Biometrics, 72(3), 877-887.

Problem definition and groups involved in methodology development

- **Problem definition**: Identifying the target population (subgroups of treatment responders) in presence of high dimensional data and survival outcome in randomized clinical trial

- Limited literature available on methodologies for this specific problem

- 2-year collaboration with different academic partners:
  - Leiden University: Prof. Hans van Houwelingen and Prof. Jelle Goeman (independent GS body)
  - Harvard School of Publich Health: Prof. LJ Wei and Prof. Tianxi Cai

Notation

Y (DFS Status, DFS Time)~ n*2 matrix of **clinical response**

G~ **treatment** (1=treated, 0=untreated)

X~ n*p matrix of **gene-expression (main effect)**

GX~ n*p matrix of **gene-treatment interaction**

Z~ n*q matrix of **clinical covariates**

On DERMA training-set: n=357; p=55 and q=11.

GS classifier: the score and the cutoff

- To build a classifier we need two main ingredients: a **score** and a **cut-off** value.

- The **score** is a continuous function of X which estimate for each patient the treatment effect (high values of the score means high probability to be GS+)

$$E(Y) = \beta_0 + X\beta_X + G\beta_G + GX\beta_{G*X}$$

$$score(X, \hat{\beta}) = E(Y|G = 1) - E(Y|G = 0) = \hat{\beta}_G + X\hat{\beta}_{G*X}$$

- a **cutoff value** to transform the score in a binary variable (GS+, GS-). The cut-off is chosen to maximize the power in the test set

Model: Weighted Logistic vs Cox

- The Cox model is the standard regression model for survival data.

- Logistic regression: probability of the events before time t_0 (weighted by the inverse of the probability to be censored).

- **PROS**: The logistic model is **more robust**. It means that the model is working even when the assumptions of the model are violated.

- **CONS**
  - results of the logistic model depends on t_0

  - the time to event is only partially used, so there is a potential loss of information respect to the Cox model.

  - observations censored before t_0 are discarded (potential loss of information)

Dimension Reduction methods

- **Principal components:** Fit a regression model using the first λ principal components.

- **PLS (Partial Least Square):** Use only the first λ factors (called PLS) explaining the covariance between [X,GX] and Y.

- **Ridge Regression:** fit a model with all the genes (main effects and interactions) and penalized [partial] likelihood (Houwelingen, 1993)

- **Random forest:** average of many decision trees (with gene-treatment interactions).

Parametrization of the interaction

Different parametrizations lead to different results

- **Classical parametrization**

$$h(t) = h_0(t) \exp\left(G\beta_G + X\beta_{X,\lambda} + GX\beta_{GX,\lambda}\right)$$

The problem is that X is more "important" than GX, because X has higher variance.

- **PG2**: prognostic effect in treated and controls.

$$h(t) = h_0(t) \exp\left(G\beta_G + GX\beta_{GX,\lambda} + (1-G)X\beta_{(1-G)X,\lambda}\right)$$

- **Two models:** one in treated and one in controls one for each model

$$h(t|G=1) = h_{01}(t) \exp\left(X\beta_{X,\lambda_1}\right); \; h(t|G=0) = h_{00}(t) \exp\left(X\beta_{X,\lambda_0}\right)$$

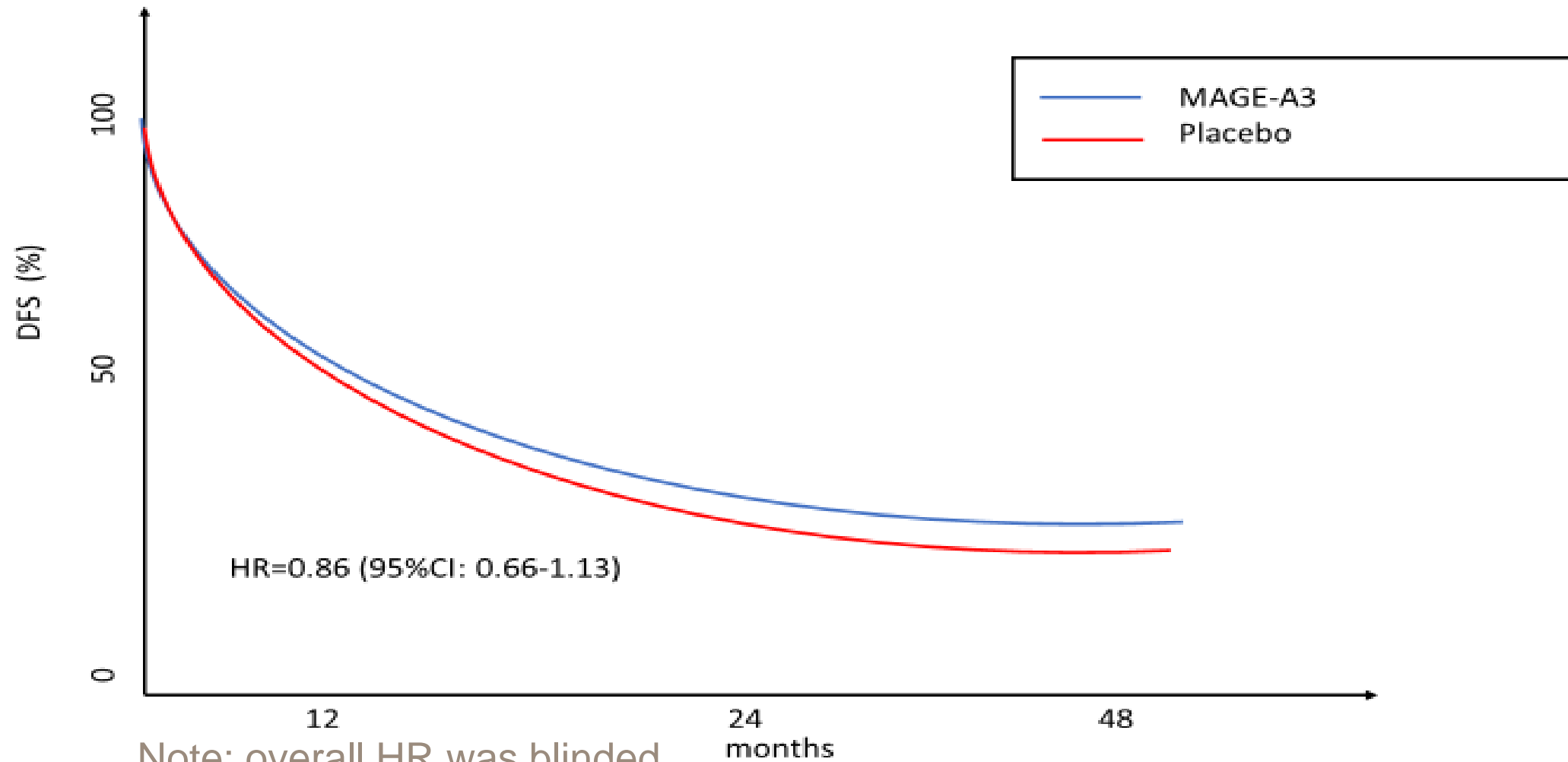this parametrization has two lambdas (λ_1, λ_0).

Results: DERMA GS

Supplementary appendix of
Dreno et al, Lancet Oncol.2018;19(7):916-929

Training set (N=366 patients)
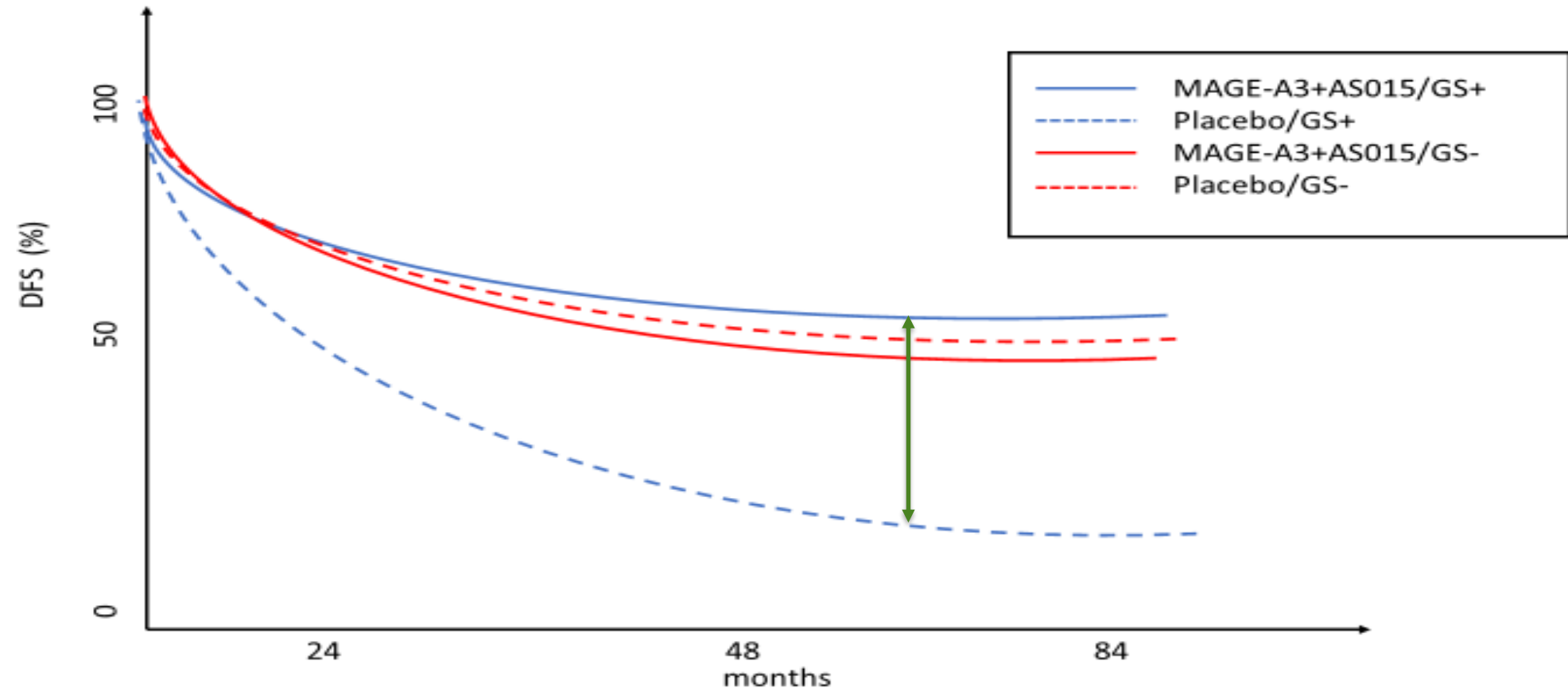
Dreno et al, Lancet Oncol.2018;19(7):916-929

HR=0.86 (95%CI: 0.66-1.13)

Note: overall HR was blinded

Classifier determination

- Different approaches (models, dimension reduction, parametrizations, tuning parameter estimation) were evaluated on simulated data and on the training-set

- **Model selected**: based on results and theoretical considerations we selected the *Ridge Cox model* with *PG2 parametrization* of the interaction and tuning parameter estimated by the *LOO cross-validated partial-likelihood*.

- **Cut-off selected**: the approach of Li et al. (2016) selected *40% of GS+ patients*
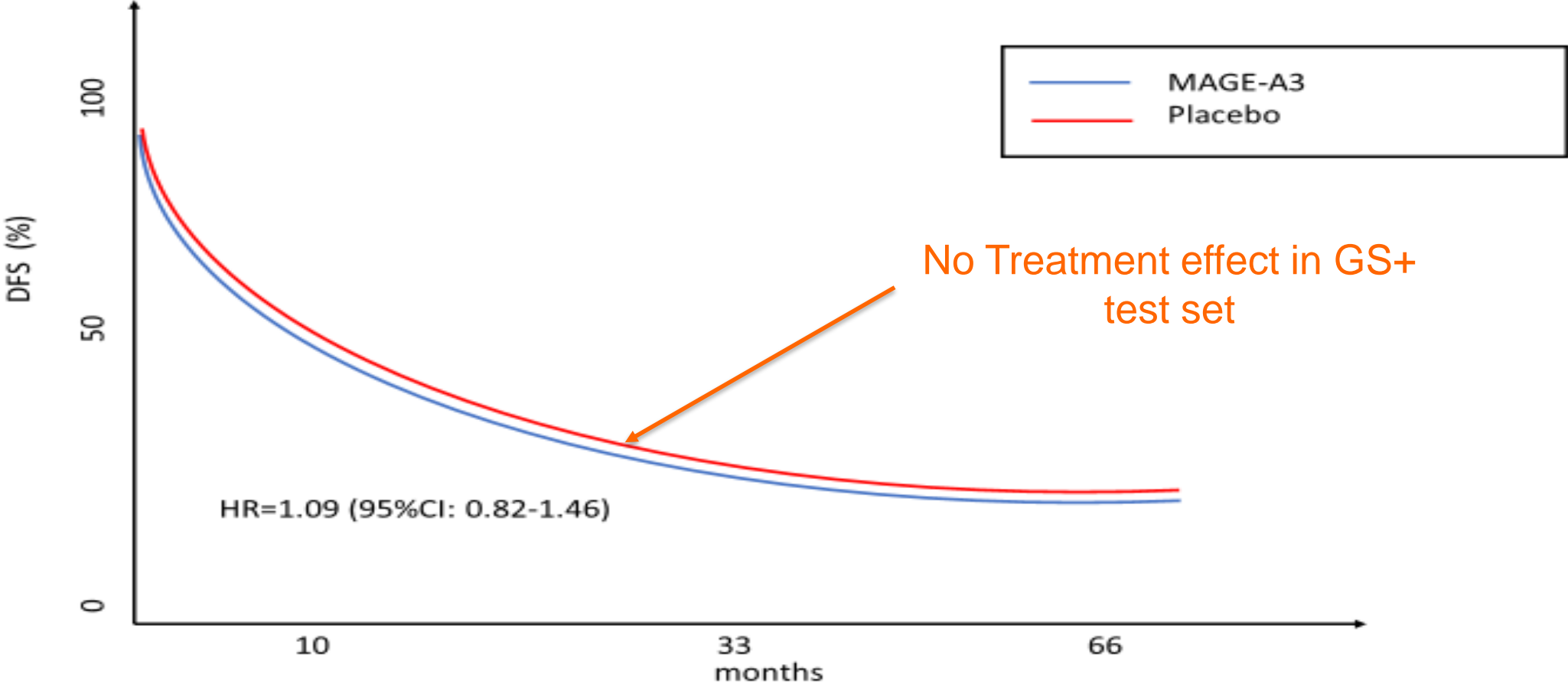
Training set by vaccination and GS

No Treatment effect in GS+ test set

HR=1.09 (95%CI: 0.82-1.46)

Adaptive signature design with futility

Biomarker clinical trials with stop for futility

Adaptive signature trials are *expensive* (measurement and validation of high-dimensional/multivariate biomarkers)

Futility: collect all the samples at baseline, but measure/validate the biomarker only if the overall treatment effect is not significant and "large enough"

- If $pv_1 \leq \alpha_1$ significant overall: biomarker not needed (H_{012} rejected)

- If $pv_1 > \alpha_1^*$ overall too small: biomarker not needed

- If $\alpha_1 < pv_1 \leq \alpha_1^*$ measure/validate the biomarker
    - if $pv_2 \leq \alpha_2$ significant GS+ (H_{012} rejected)

Responders and Non-responders assumption

The overall treatment effect is

$$E(U_1) = E(\bar{y}_A - \bar{y}_B) = \pi \Delta_R + (1 - \pi)\Delta_{NR}$$

where π is the proportion of Responders and $\Delta_{NR} = \beta \Delta_R$ is the treatment effect in non-responders.

So the treatment effect in GS+ subgroup is

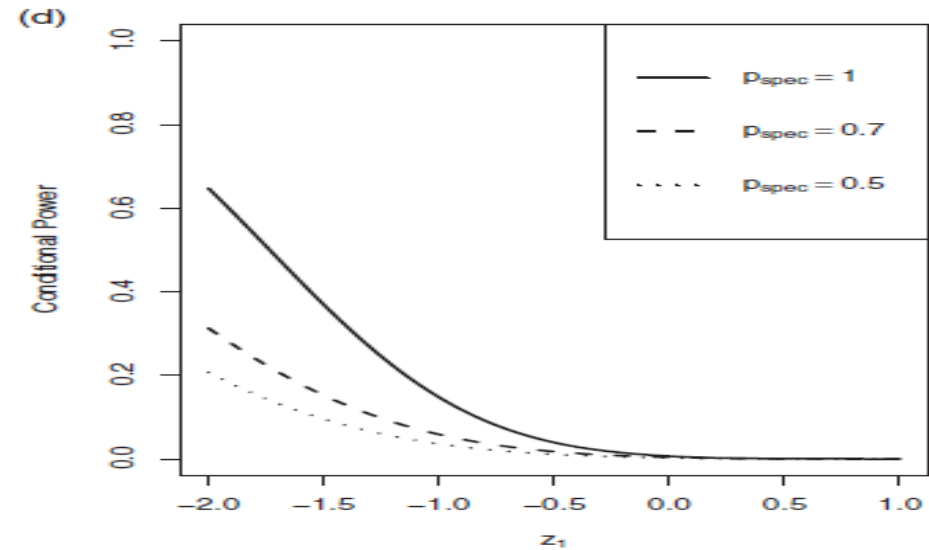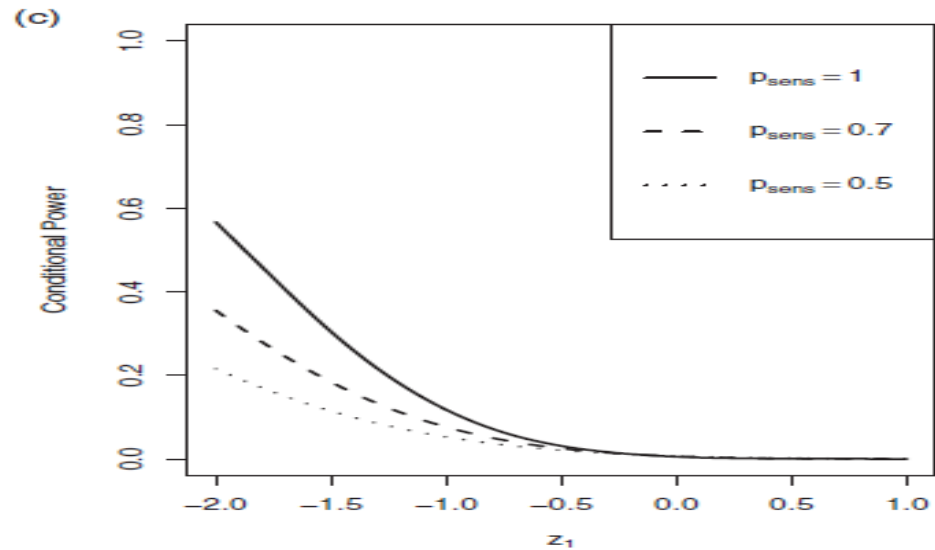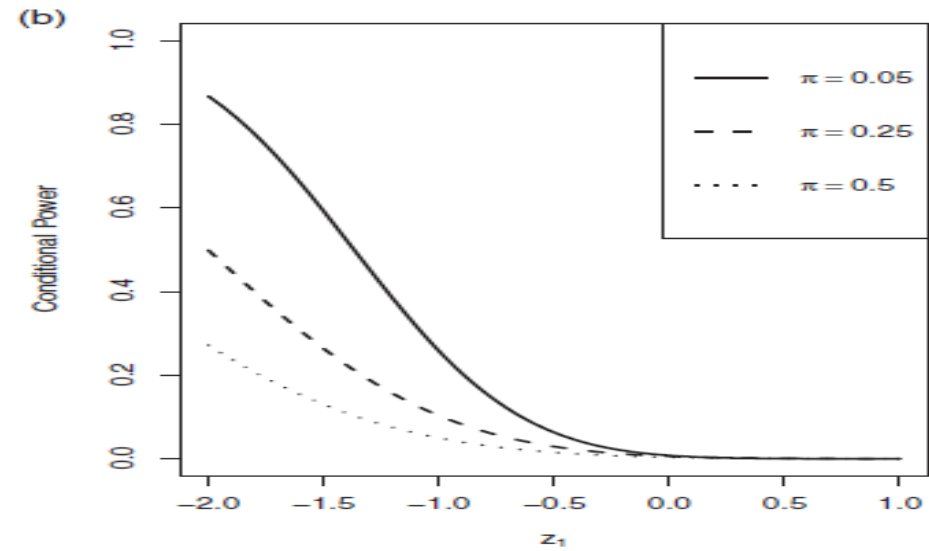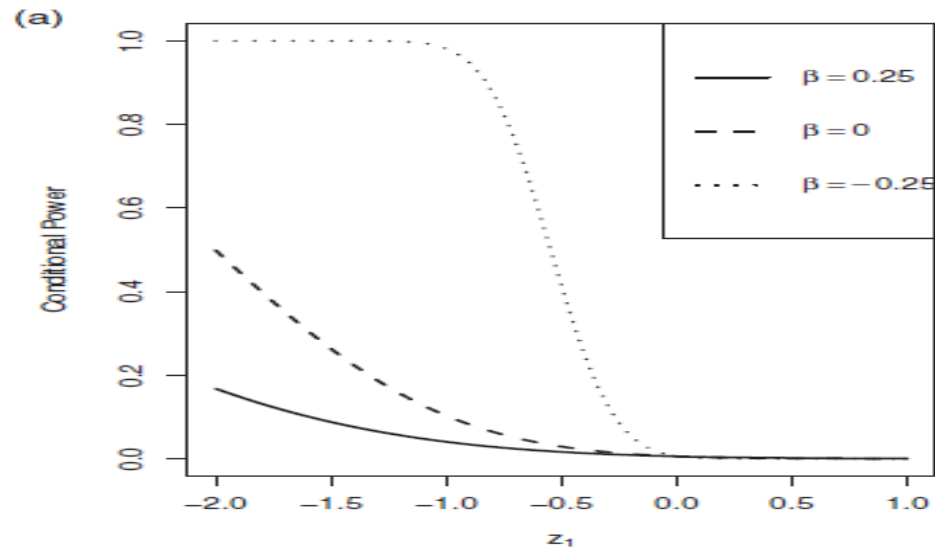$$E(U_2) = \Delta_R [PPV + (1 - PPV)\beta]$$

where the Positive Predictive Value (PPV) is

$$PPV = \frac{\pi p_{sens}}{p_+}$$

where p_+ is the proportion of subjects in the GS+ and p_{sens} is the sensitivity of the GS classifier to identify responders.

CP is large when
- treatment is null/harmful in non-responders,
- small proportion of responders
- Good GS+ classifier (high sensitivity/ specificity).

Overall treatment effect "large enough"

Biomarker measured/validated only if $\alpha_1 < pv_1 \leq \alpha_1^*$

$\gamma = 0.2, \ \alpha = 0.025, \ \alpha_1 = 0.02, \ p_{test} = 0.5, \ p_{spec} = 0.9 \quad \beta = 0$

Assumption: no treatment effect in non-responders

π	p_{sens}	p_+	α_2^{SD}	$\alpha_1^*(CP(\hat{\theta}))$
0.05	1.0	0.145	0.006	0.176
0.05	0.7	0.130	0.005	0.104
0.05	0.5	0.120	0.005	0.045
0.25	1.0	0.325	0.006	0.080
0.25	0.7	0.250	0.006	0.040
0.25	0.5	0.200	0.006	0.014
0.50	1.0	0.550	0.007	0.033
0.50	0.7	0.400	0.006	0.013
0.50	0.5	0.300	0.006	0.003

No need to measure the biomarker if overall pvalue>0.2 (futility based on overall results).

Adaptive signature not useful when proportion of responders is large and/or when the quality (sens./spec.) of the GS classifier is low.

Futility based on the training-set biomarker results

- If $\alpha_1 < pv_1 \leq \alpha_1^*$ measure the biomarker in the training-set

- validate/measure the biomarker in test-set only if biomarker results are promising

- Futility can be based on
  - Conditional Power (DERMA)
  - interaction between the high-dimensional biomarker and the treatment (Callegaro et al, Biom. Journal 2017 59(4), 672-684.).
    - not necessary to build the GS classifier

Conclusions

We showed

a real implementation of Adaptive Signature Design (ASD)

- multivariate qRT-PCR (55 genes selected in Phase II).

- GS not working on the test set
  - *no treatment effect: positive in training-set and negative in test-set*

statistical challenges to build the GS

- high-dimensional data; parametrization of the interaction; estimation of the tuning parameter; cut-off determination…

ASD with futility based on

- overall results

- training-set biomarker results

Conclusions

ASD is useful when

i)   proportion of responders is small (but not too small)

ii)   good GS classifier (high sensitivity and specificity)

More chance to have a good GS classifier if the biomarker is

- **"validated"** (good control of non-biological variability)

- **biologically-informed**
  - biomarker research/exploration in early-phase trials
  - *good* data vs *big* data
  - *expertise* vs *black box*

Main References

- Callegaro, A (2019). Futility for subgroup analyses in the adaptive signature design. *SMMR* ;28(3):953-961.

- Callegaro, A et al. (2017). Testing interaction between treatment and high-dimensional covariates in randomized clinical trials. *Biometrical Journal*, 59(4), 672-684.

- Dreno, B et al. (2018). MAGE-A3 immunotherapeutic as adjuvant therapy for patients with resected, MAGE-A3-positive, stage III melanoma (DERMA): a double-blind, randomised, placebo-controlled, phase 3 trial. *The Lancet Oncology*, 19(7), 916-929.

- Freidlin B and Simon R (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res*; 11: 7872–7878.

- Li, J et al. (2016). A predictive enrichment procedure to identify potential responders to a new therapy for randomized, comparative controlled clinical studies. *Biometrics*, 72(3), 877-887.

- Ulloa-Montoya, F et al. (2013). Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *J Clin Oncol*, 31(19), 2388-2395.

Thank you