Evaluating methods for biomarker cut-off detection: a comparative study in small and unbalanced samples

H. Hadjur*, J. Geronimi, S. Guilleminot, A. Serra, L. Daniells, M-K. Riviere, G. Saint-Hilary, P. Mozgunov

***Department of Statistical Methodology, Saryga**

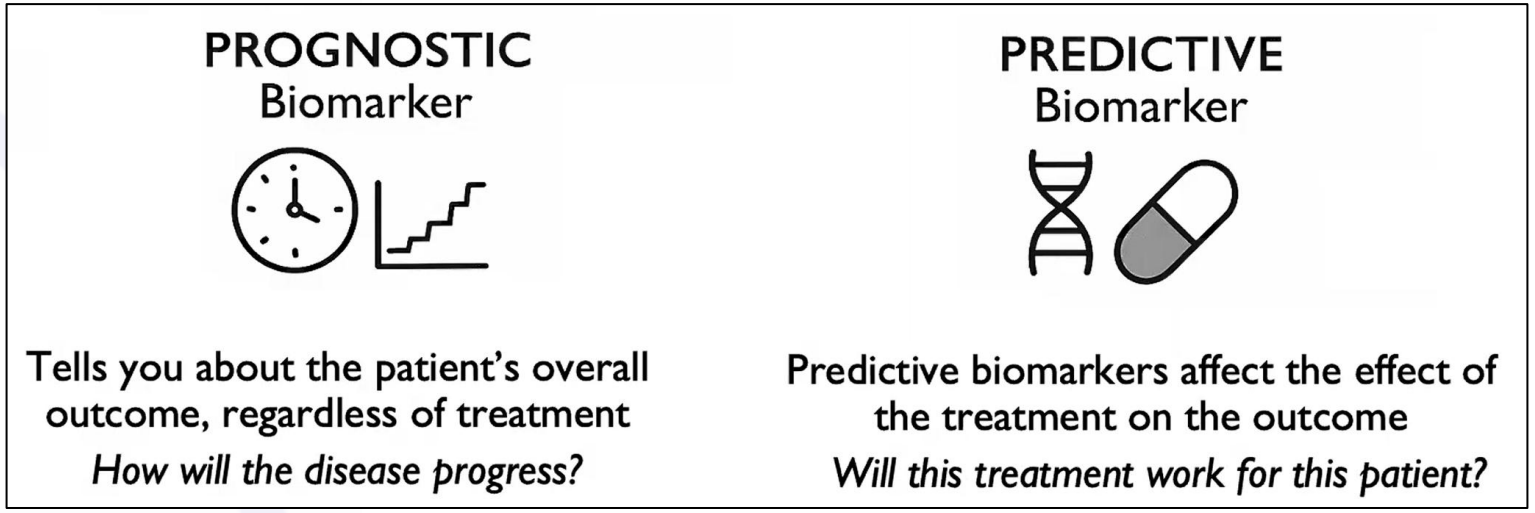
Biomarkers SIG and Treatment Effect Heterogeneity SIG

PSI 2026 CONFERENCE

SERVIER moved by you

SARYGA STATISTICS & METHODOLOGY

Background – Predictive biomarker and cut-off



- **Predictive biomarkers** identify patients most likely to benefit from a therapy

- In that context, a biomarker cut-off enables the **assignment of treatments to subpopulations** so that **response is optimised**

SARYGA
STATISTICS & METHODOLOGY

Background – Cut-offs shape clinical development

- A more **inclusive cut-off increases access but may dilute the observed treatment effect**; a more **restrictive cut-off strengthens the signal but excludes patients who might still benefit**.

- Because of these trade-offs, **regulators and HTA bodies expect cut-offs to be robust, reproducible and clinically interpretable**, in addition to statistically convenient.

(FDA) U. In Vitro Companion Diagnostic Devices Guidance for Industry and Food and Drug Administration Staff. Guidance for Industry and FDA Staff; 2014. Available as FDA Document 81309.

SARYGA
STATISTICS & METHODOLOGY

Approach

- We explore a number of methods to **identify cut-off values for a continuous predictive biomarker**
- We focus on **small and unbalanced sample (40:20)** and explore both **one-arm and two-arm methods**
- We aim to:
  - Give insight into the **trade-offs between different cut-off identification methods**
  - Provide **practical guidance for choosing the right approach** in challenging clinical situations

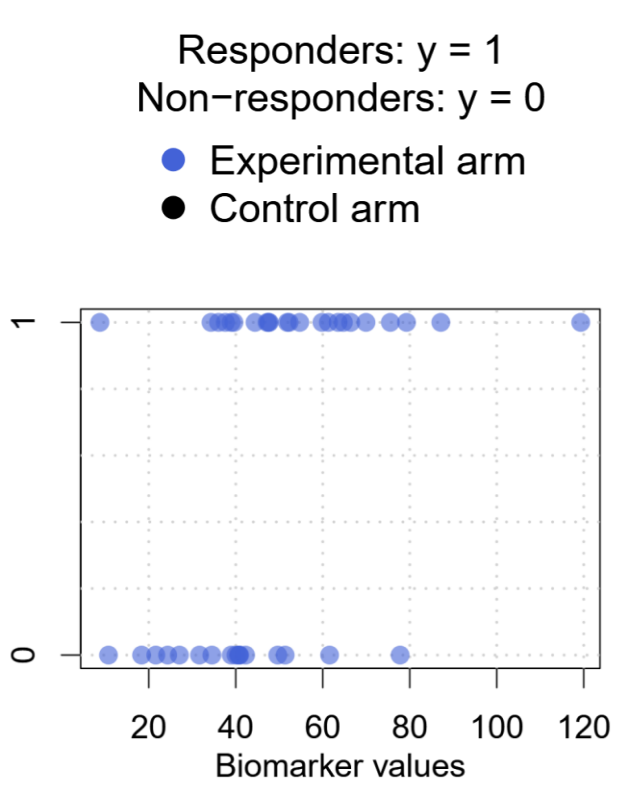
SARYGA
STATISTICS & METHODOLOGY

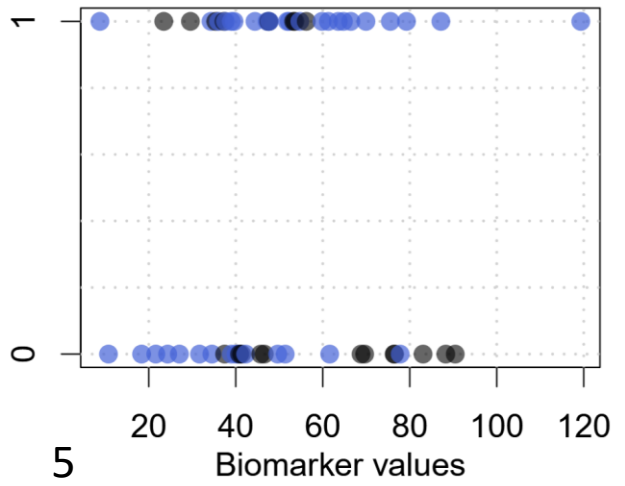
Cut-off identification methods

One-arm vs two-arm

- **One-arm** methods only use the experimental arm
- The **do not distinguish predictive from prognostic effects**
- They **do not estimate treatment effect**
- They **optimise discrimination metrics**

- **Two-arm** methods use both arms
- They can model the **treatment-biomarker interaction**, thus the **treatment effect**
- **Prognostic and predictive effects can be estimated separately**





Cut-off identification methods

One-arm methods

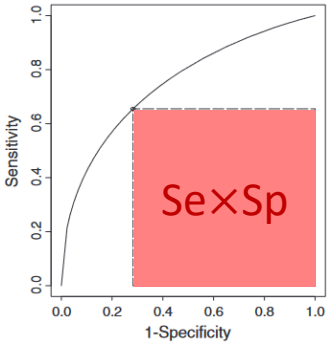
Note: the support of x is the set of possible biomarker values

- **Product of sensitivity and specificity** (Se×Sp)

The tested model is: assign positive responses for values above the cut-off point and negative response for values below the cut-off point.

$$c = \underset{x}{\mathbf{argmax}}(Se(x) \times Sp(x))$$

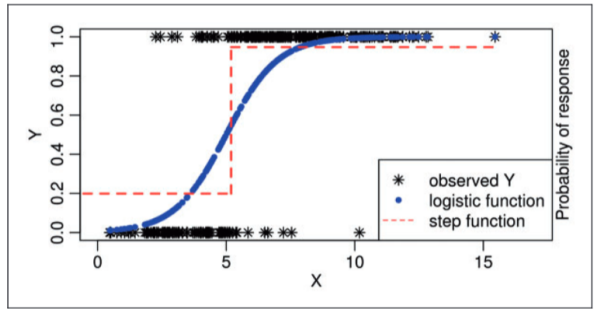
The area of the largest rectangle defined under the ROC curve



- **Optimal step function based on ordinary least squares*** (SF-OLS)

$$c = argmin_c \sum_x (P(Y = 1|X = x) - SF(x, c))^2$$

The "step value" of the step function obtained by the ordinary least squares method



* We use a frequentist approach inspired from: Vradi E, Jaki T, Vonk R, Brannath W. A Bayesian model to estimate the cutoff and the clinical utility of a biomarker assay. Statistical Methods in Medical Research 2019 doi: 10.1177/0962280218784778

SARYGA
STATISTICS & METHODOLOGY

Cut-off identification methods

Two-arm methods

- **Predictive biomarker graphical approach** (PRIME*): fit one logistic regression model $g(\beta_1 T + \beta_2 X + \beta_3 TX)$ and find the point where the 80% lower confidence interval of the treatment effect crosses 0

  Optimal cutoff: $\boldsymbol{c = \underset{x}{\mathbf{argmin}}(lower\_CI(|\hat{y}_{exp} - \hat{y}_{ctrl}|))}$

- **Optimal step function based on ordinary least squares**

Optimal cut-off: "step value" of the step function obtained by the ordinary least squares methods

"OLS-3-steps": one below-threshold step with both arms, two above-threshold steps (one for each arm)

- **Dichotomised Logistic Modelling** (DLM): model-based approach where a **logistic regression is fitted for dichotomised datasets** (at each potential cut-off value)

  Optimal cutoff: $\boldsymbol{c = \underset{x}{\mathbf{argmin}}(pvalue_{logistic\ model})}$

*Reference for the "point estimate" variant: D'Angelo G, Tian X, Deng C, Zhou X. Predictive Biomarker Graphical Approach (PRIME) for Precision Medicine. Pharm Stat. 2026 doi: 10.1002/pst.70094. PMID: 42050979 – We introduce the two CI variations PRIME80% and PRIME50%

SARYGA
STATISTICS & METHODOLOGY

Data setting

X	T	Y
16.4	0	0
39.3	1	1
57.1	1	1
68.3	0	1

Example of data

- **X is a continuous biomarker**
- X follows a Uniform or Gamma distribution

- **T is a binary treatment indicator**
- $T = 0$ defines the control arm and $T = 1$ defines the experimental arm

- **Y is a binary outcome** of the primary interest in the trial
- Y follows a Bernoulli distribution with probability p
- $Y = 0$ refers to a negative response and $Y = 1$ refers to a positive response to treatment

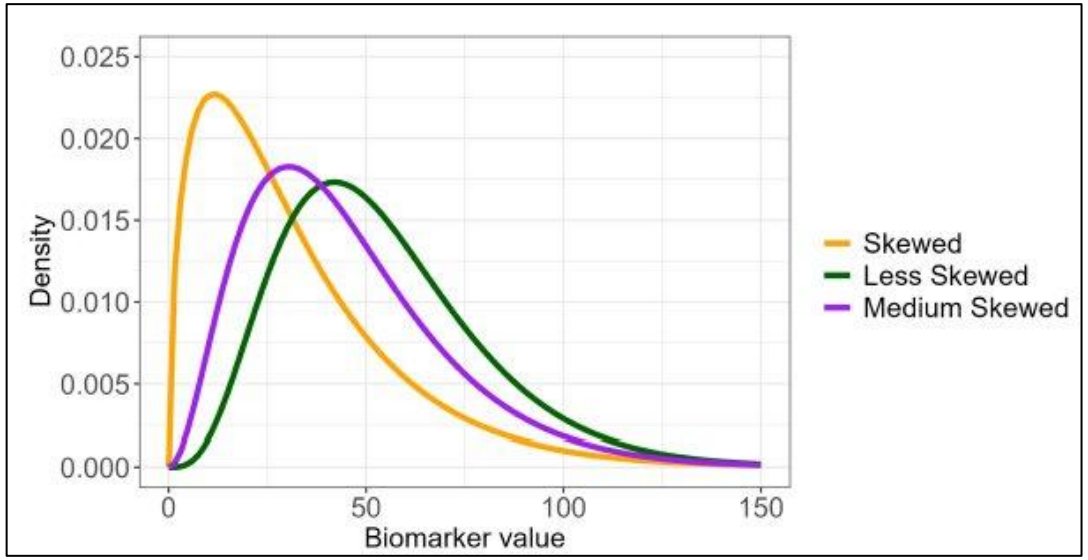
- A logistic relationship is used to map X and T to p:

$$p = \frac{\exp(b_0 + b_1 T + b_2 X \times T + b_3 X)}{1 + \exp(b_0 + b_1 T + b_2 X \times T + b_3 X)}$$

SARYGA
STATISTICS & METHODOLOGY

Simulation setting

- 5 methods (2 one-arm, 3 two-arm)
- 6 response-biomarker scenarios with varying predictive and prognostic signal
- R = 5000 simulations per scenario
- Less-skewed Gamma distribution (main setting)
- Sample size: 40 experimental, 20 controls

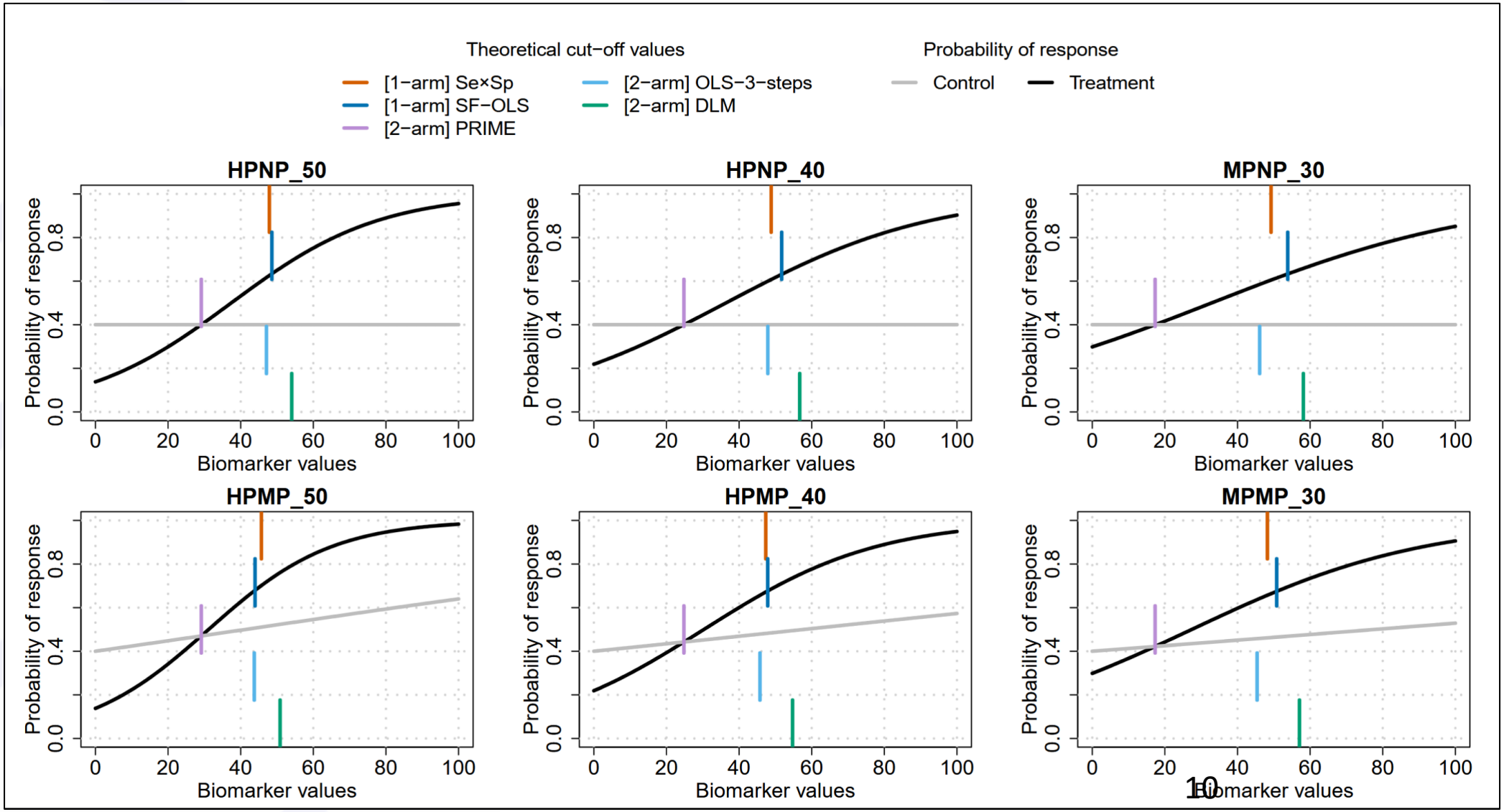


Considered Gamma biomarker distributions

SARYGA
STATISTICS & METHODOLOGY

Simulation setting

Scenarios



SARYGA
STATISTICS & METHODOLOGY

Simulation setting

Two-step approach

For each simulation sample i:

- **Biomarker predictiveness is assessed** using the AKSA method (Serra et al., 2025)

Reminder of AKSA's general idea: estimate the "global trend" of the difference in responses between treatment and control over the range of biomarker values while accounting for uncertainty around this difference

- **Only when predictiveness is established: the cut-off values $\widehat{c_i}$ are identified using the methods**. A cut-off value always belongs to the dataset's list of biomarker values

Serra A, et al. A Novel Approach to Assess the Predictiveness of a Continuous Biomarker in Early Phases of Drug Development. Statistics in Medicine 2025; 44(5): e70026. doi: https://doi.org/10.1002/sim.70026

SARYGA
STATISTICS & METHODOLOGY

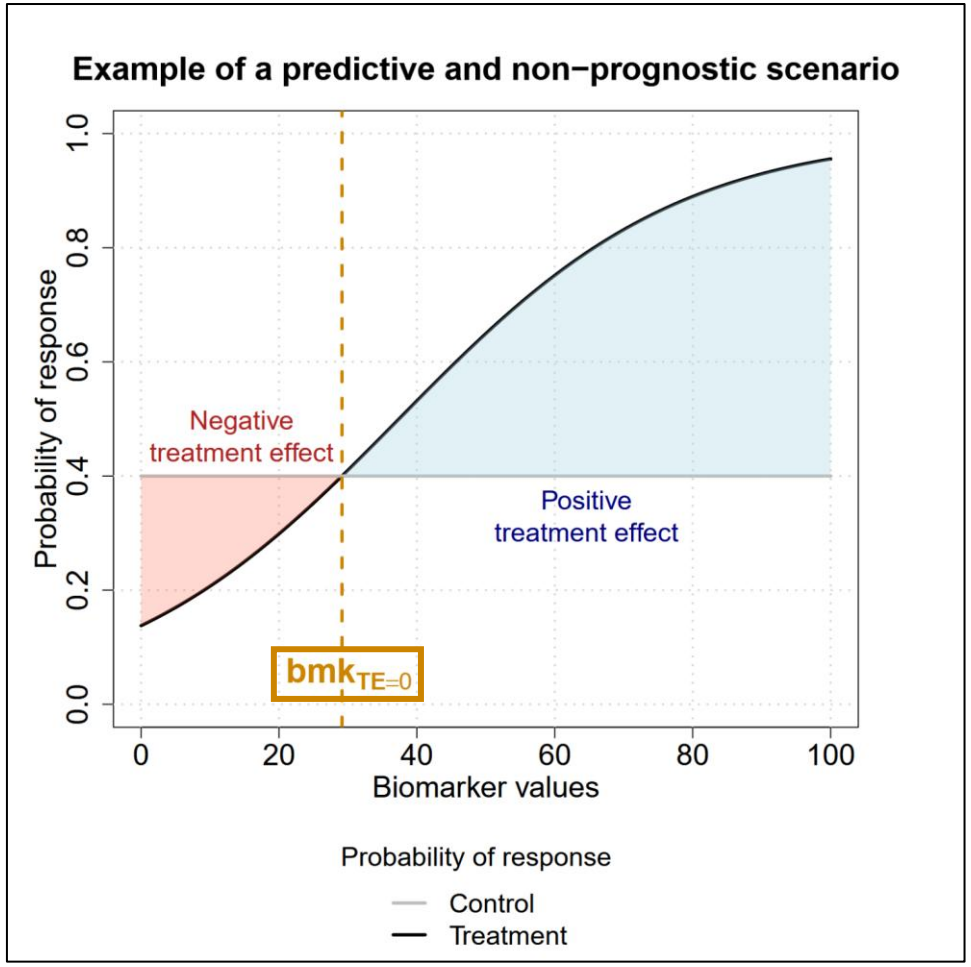
Simulation setting

Metrics of interest

- **Bias**: $\frac{1}{R}\sum_{i=1}^{R}(\widehat{c_i} - c_{true})$, where c_{true} is the analytical cut-off
- **Standard deviation** of the simulation cut-off values $\widehat{c_i}$
- **Root mean squared error** ($\sqrt{\text{MSE}}$): $Var(\hat{c}) + Bias(\hat{c})^2$

- $\boldsymbol{P(c > bmk_{TE=0})}$: probability that the identified biomarker cut-off value is greater than $\boldsymbol{bmk_{TE=0}}$, which is the biomarker value from the distribution associated with the treatment effect crossing 0 (see plot).

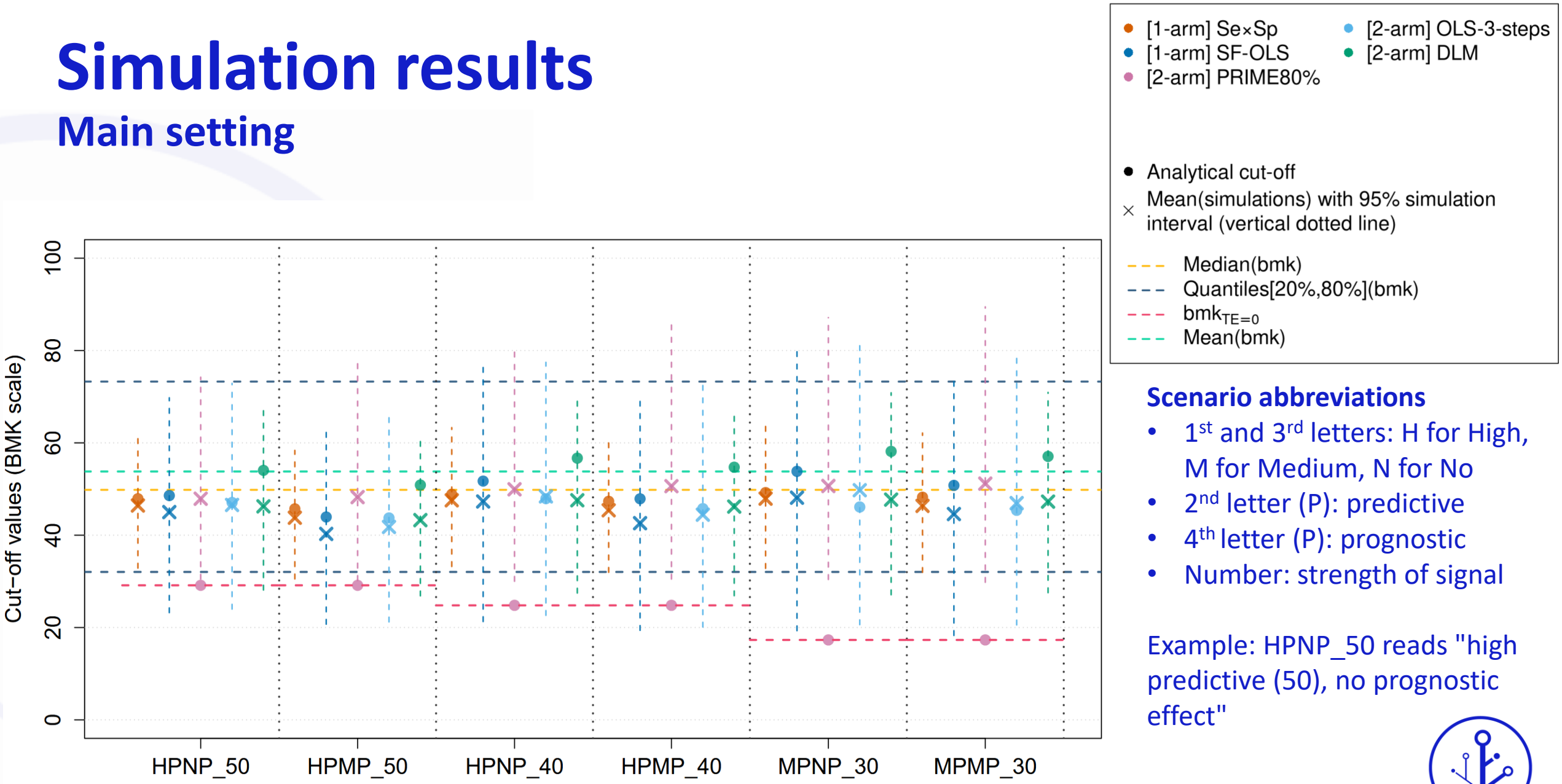
- **Mean BMK+ population proportion**: the mean proportion of the BMK+ population across simulation runs. It quantifies whether the cut-off identification methods favours broad or narrow definitions of the BMK+ subgroup.



SARYGA
STATISTICS & METHODOLOGY

Simulation results

Main setting



Scenario abbreviations

- 1st and 3rd letters: H for High, M for Medium, N for No
- 2nd letter (P): predictive
- 4th letter (P): prognostic
- Number: strength of signal

Example: HPNP_50 reads "high predictive (50), no prognostic effect"

SARYGA
STATISTICS & METHODOLOGY

Simulation results

Main setting

- Simulation results show how well each method targets its analytical cut-off point

Method	1-arm		2-arm		
Performance metric	**Se×Sp**	**SF-OLS**	**PRIME80%**	**OLS-3-steps**	**DLM**
Bias	**-1.6**	**-4.8**	18.7	**0.3**	-8.9
Std. dev.	**7.6**	**13.6**	11.2	14	**10.8**
√MSE	**7.7**	**14.4**	21.9	**14.1**	**14**
$\boldsymbol{P(c > bmk_{TE=0})}$	1	0.93	0.95	0.94	0.98
BMK+ population proportion	0.55	0.58	0.63	0.56	0.56

- One-arm methods
  - Best under our setting = **Product of sensitivity and specificity (Se×Sp)**
  - Worst under our setting = **Optimal step function based on ordinary least squares (SF-OLS)**

- Two-arm methods
  - Best under our setting = **OLS-3-steps the least biased, DLM the most consistent**
  - **PRIME is the most clinically interpretable:** it picks the point above which there is a positive treatment effect

SARYGA
STATISTICS & METHODOLOGY

Next steps

- Stay tuned: our publication is currently under revision!
- Currently, we use logistic probability of response and logistic-based models for some cut-off identification methods → next, **add non-logistic scenarios**
- Consider **ML-based** methods for cut-off identification
- We are currently working on the analysis of **treatment effect estimation** methods

Thank you!

hugo.hadjur@saryga.com

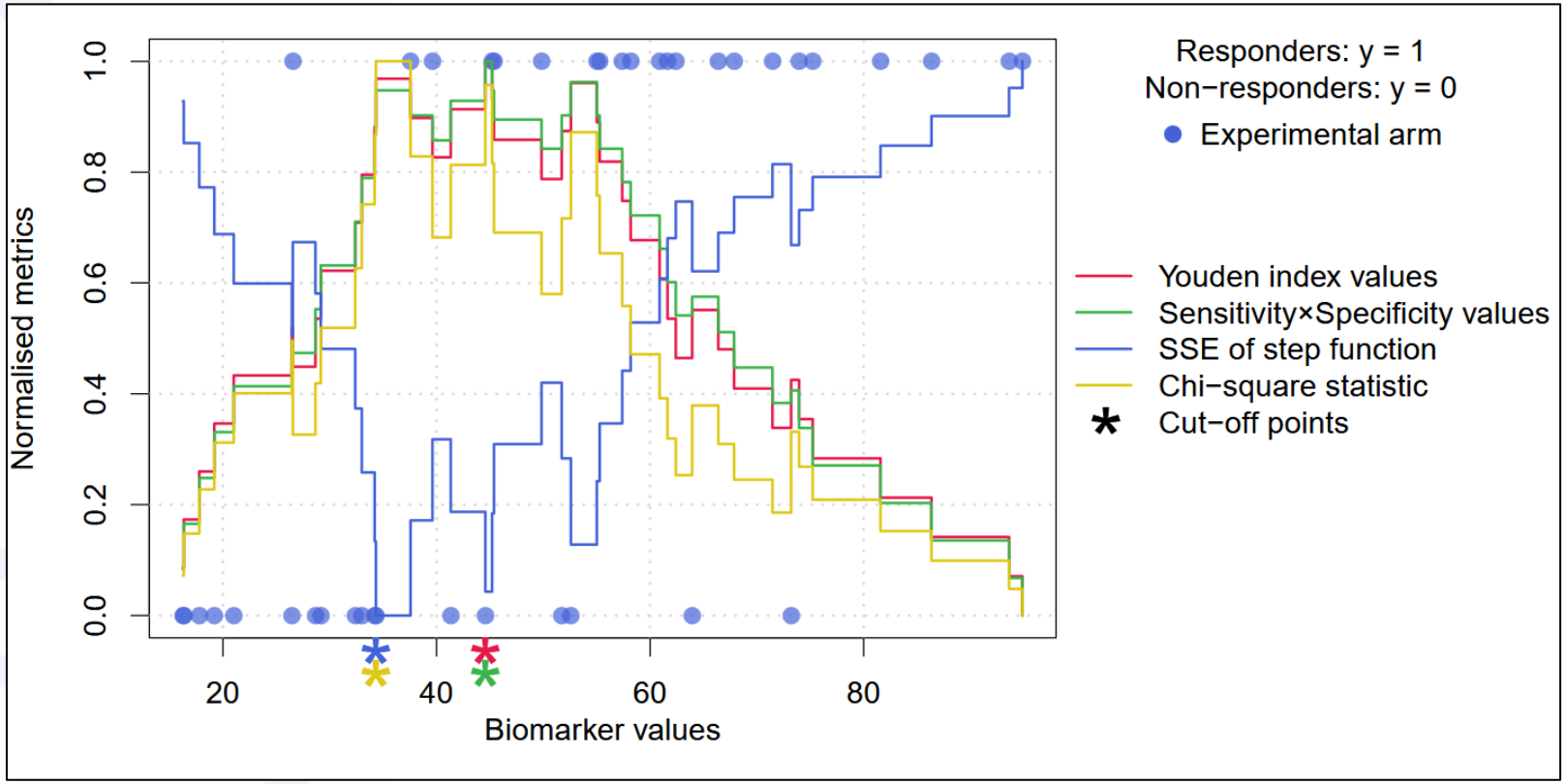
SARYGA
STATISTICS & METHODOLOGY

Backup

SARYGA
STATISTICS & METHODOLOGY

Cut-off identification methods

One-arm methods – Example



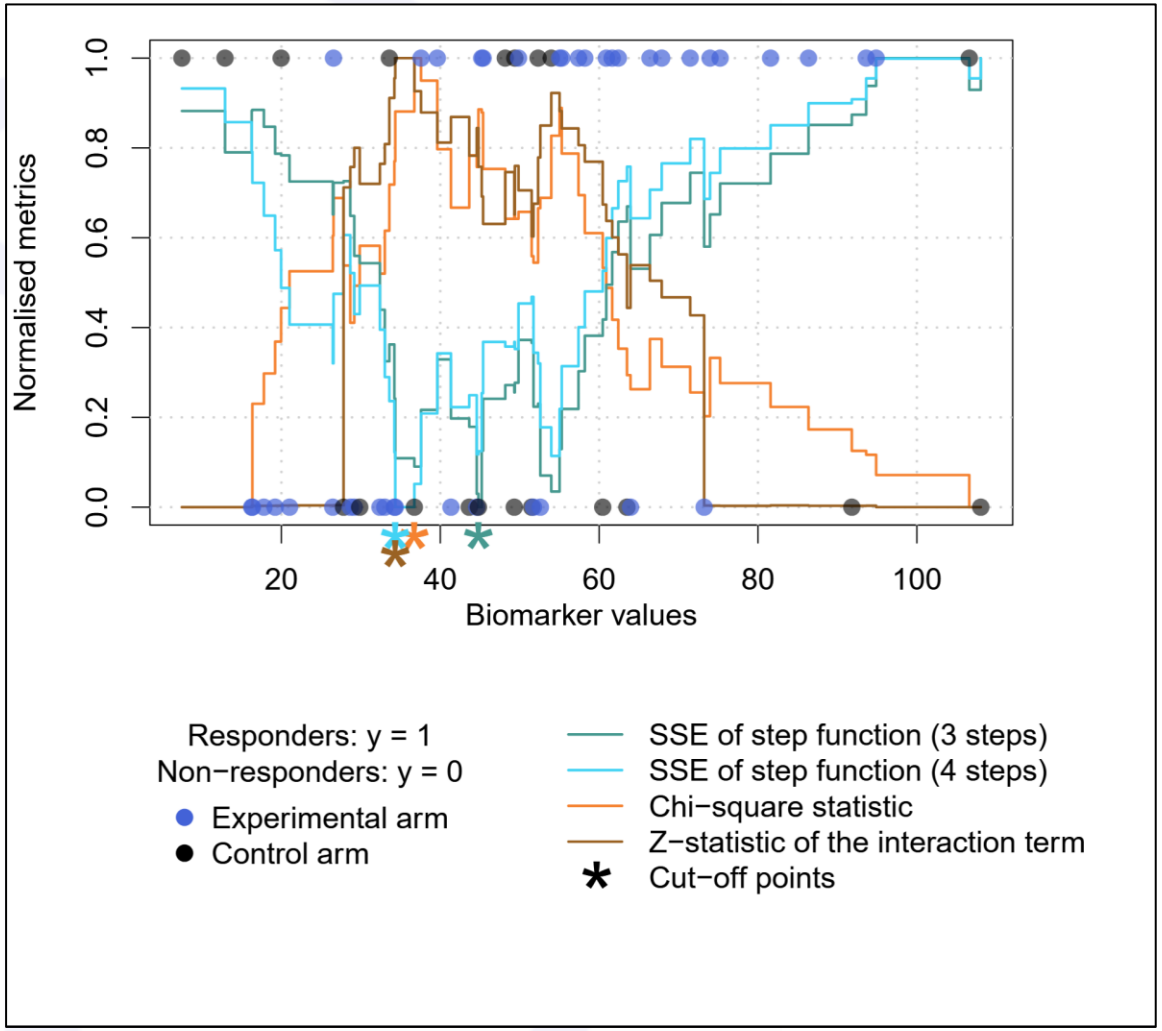
Application of the one-arm methods on an example 40-20 dataset

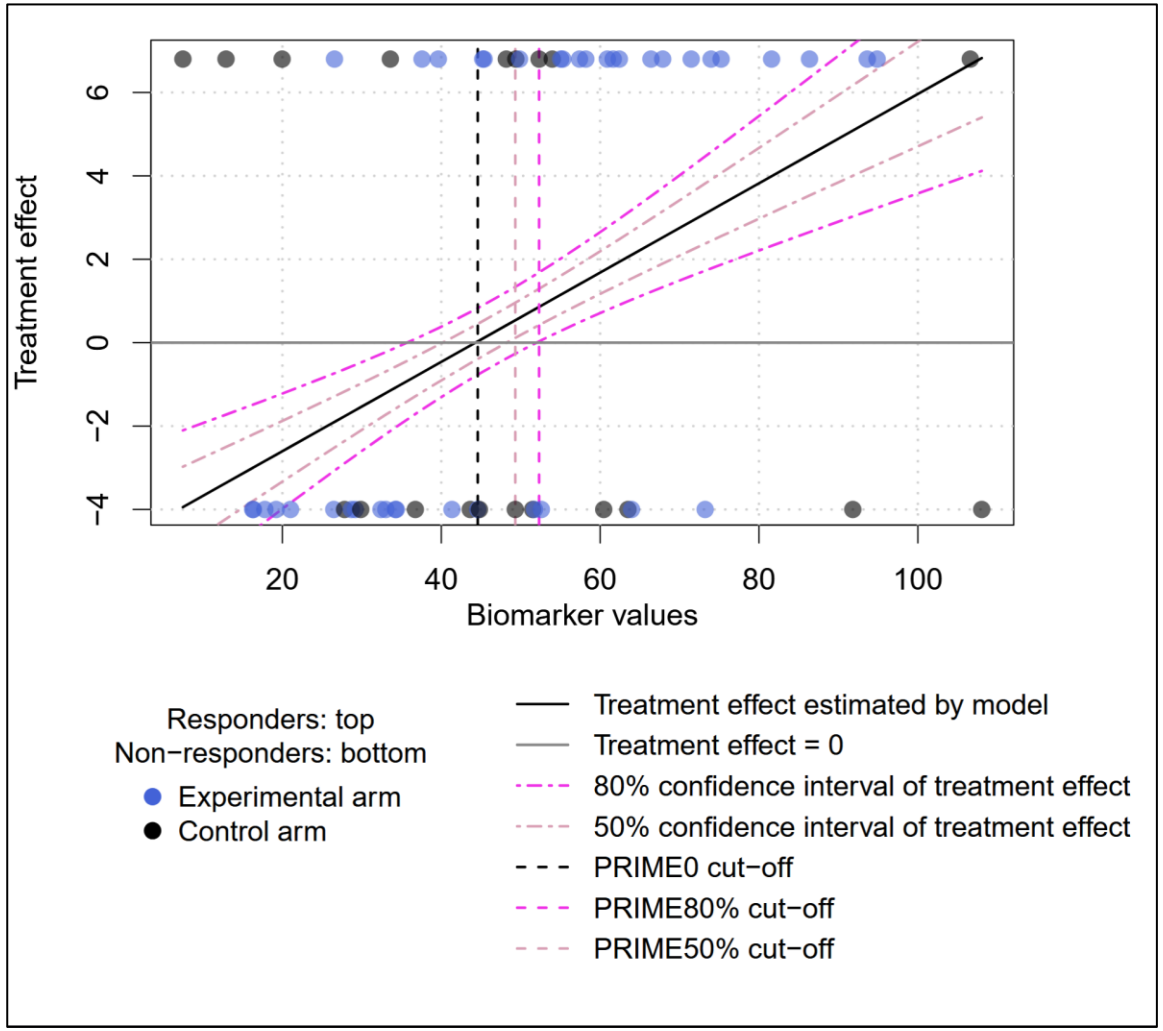
Cut-off values are selected from the data points

SARYGA
STATISTICS & METHODOLOGY

Cut-off identification methods

Two-arm methods – Example



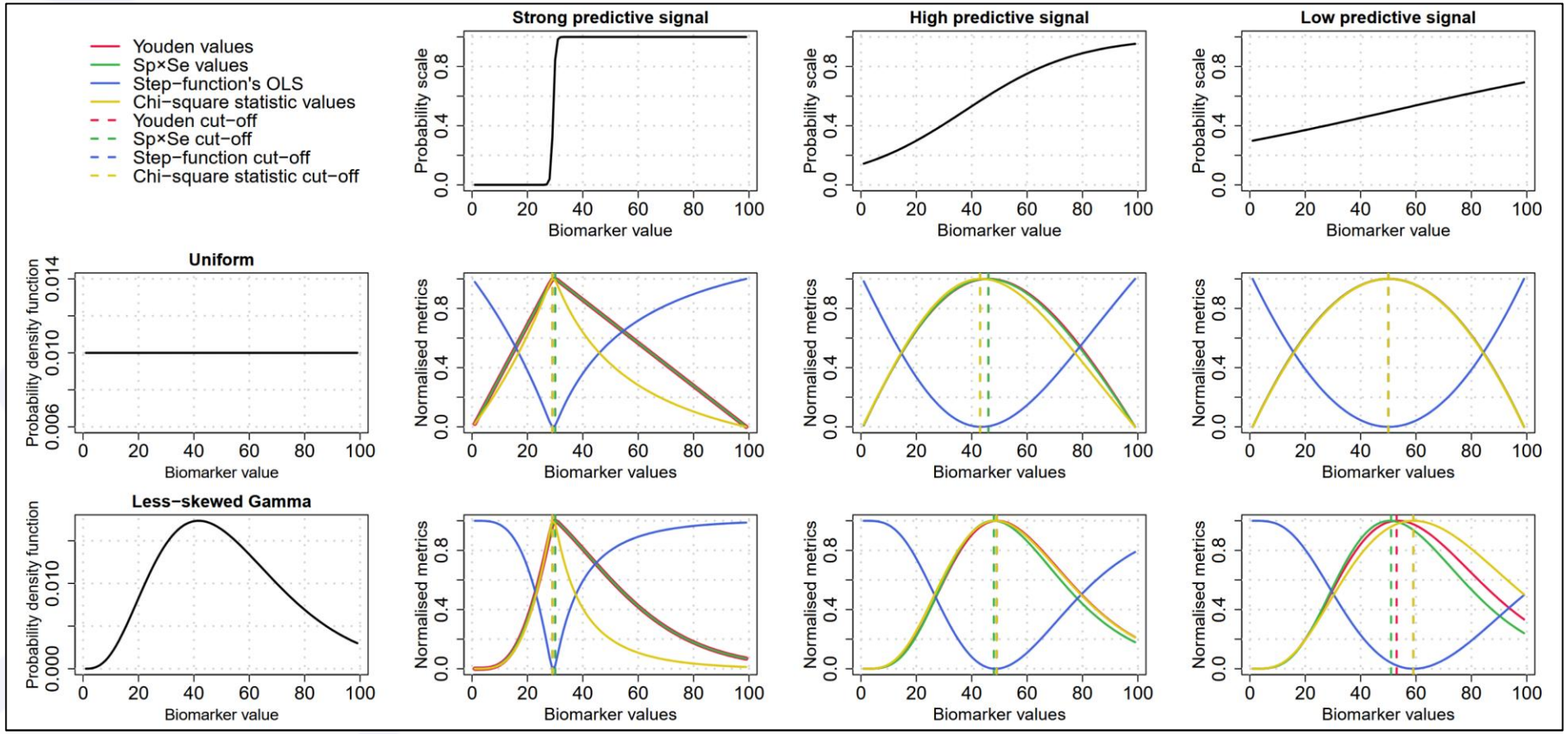


Application of the two-arm methods on an example 40-20 dataset

SARYGA
STATISTICS & METHODOLOGY

Cut-off identification methods

One-arm methods – Analytical values



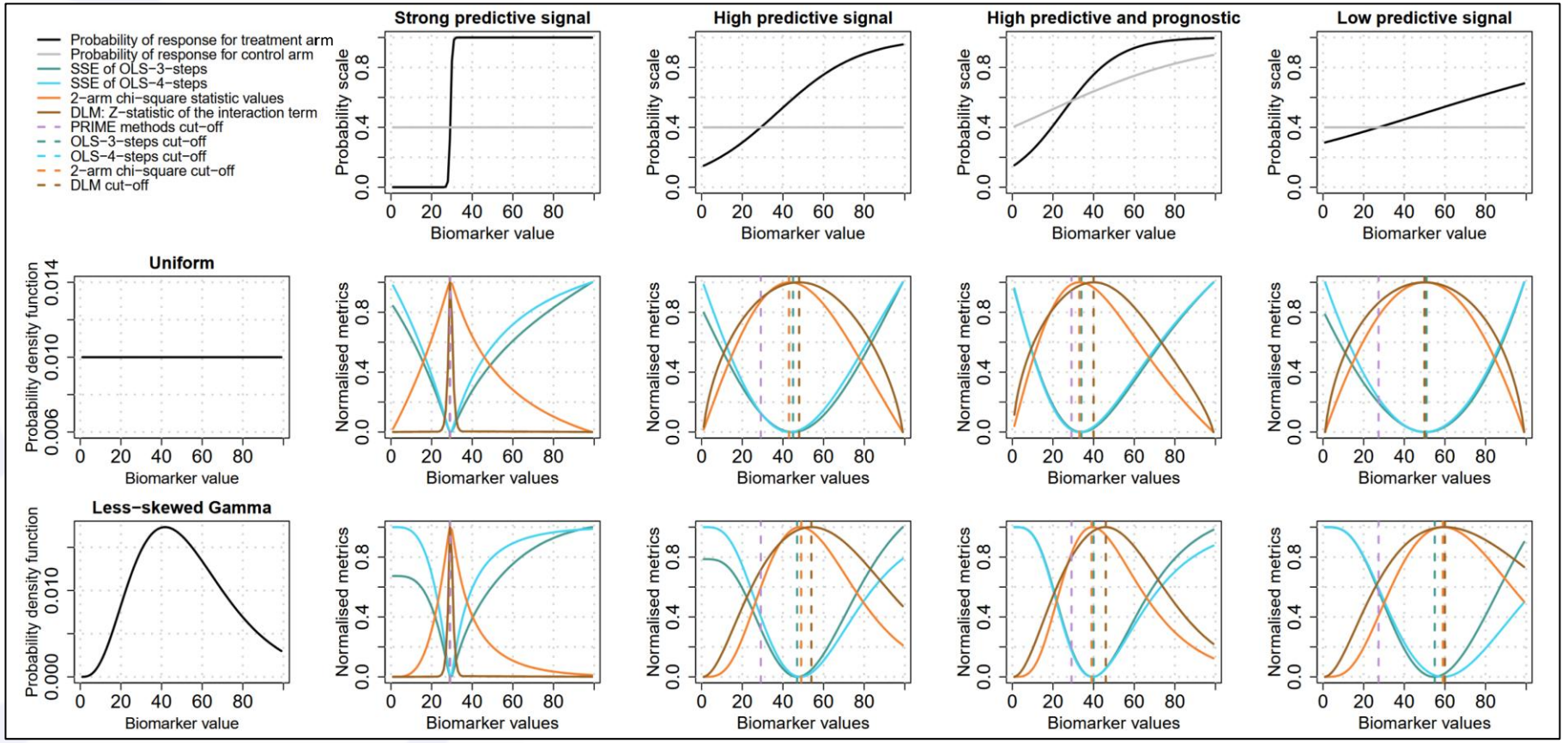
Analytical values and cut-off points for the one-arm methods given six pairs of biomarker distribution and response probability

- Given distributions for the biomarker and the response probability, analytical values can be obtained
- SF-OLS and 1-arm-Max-χ^2 give identical analytical cut-off points given all the considered scenarios
- The choice of the method does not impact the analytical cut-off when the predictive signal is strong

SARYGA
STATISTICS & METHODOLOGY

Cut-off identification methods

Two-arm methods – Analytical values



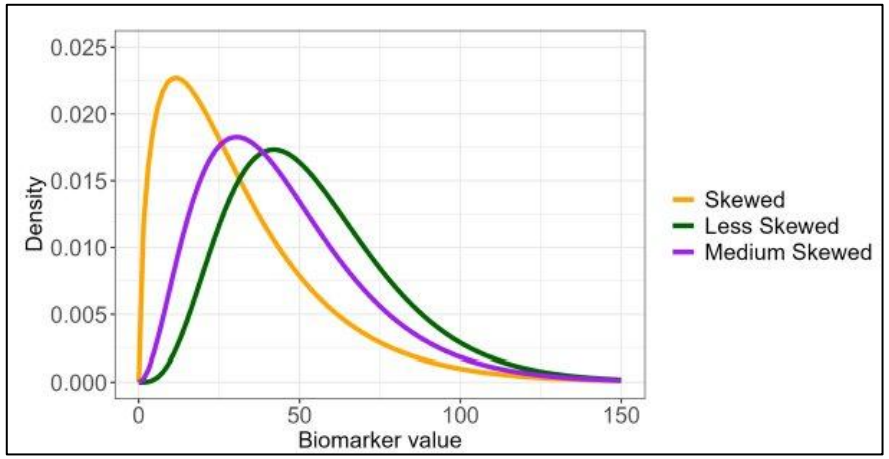
Analytical values and cut-off points for the two-arm methods given eight pairs of biomarker distribution and response probability

Analytical and data-simulated cut-off points are used to compute the performance metrics

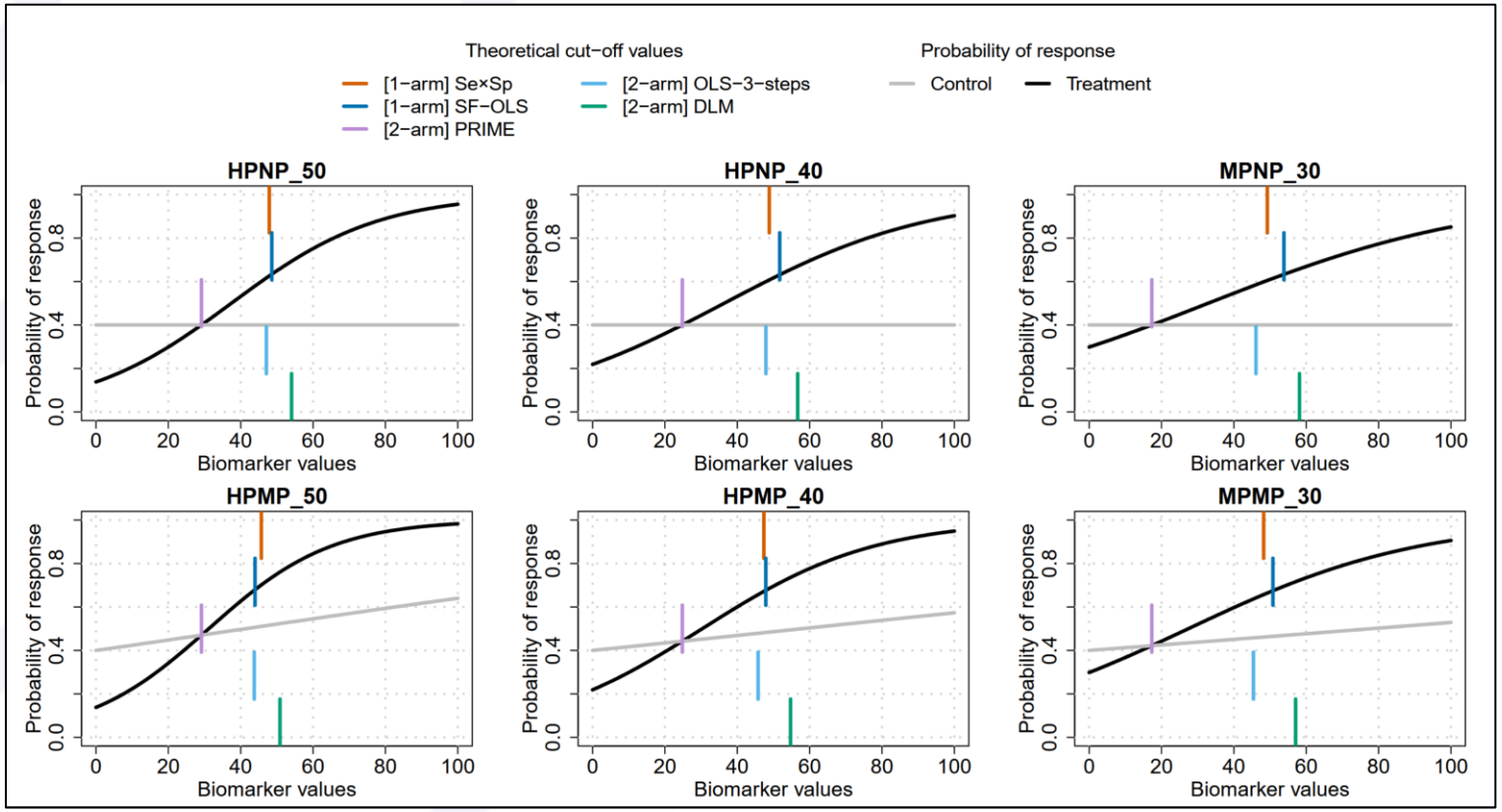
SARYGA
STATISTICS & METHODOLOGY

Simulation setting

- 10 methods (3 one-arm, 7 two-arm)
- 6 scenarios and R = 5000 simulations per scenario
- Less-skewed Gamma distribution (main setting)
- Sample size: 40 experimental, 20 controls



Biomarker distributions



Scenarios for the probability of response and their associated analytical cut-off points given a "less-skewed" biomarker distribution

Scenario abbreviations

- 1st and 3rd letters: H for High, M for Medium, N for No
- 2nd letter (P): predictive
- 4th letter (P): prognostic
- Number: strength of signal

Example: HPNP_50 reads "high predictive (50), no prognostic effect"

SARYGA
STATISTICS & METHODOLOGY

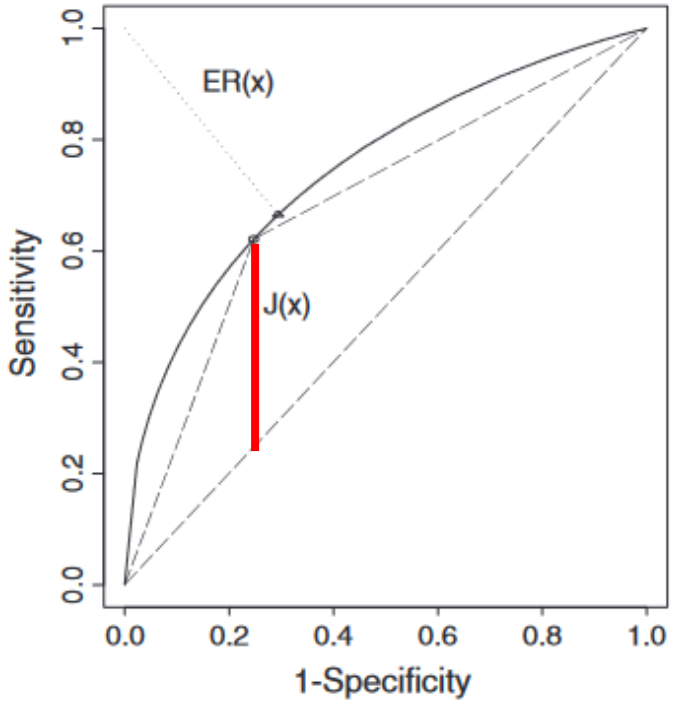
Youden index
Youden 1950 & Fluss 2005

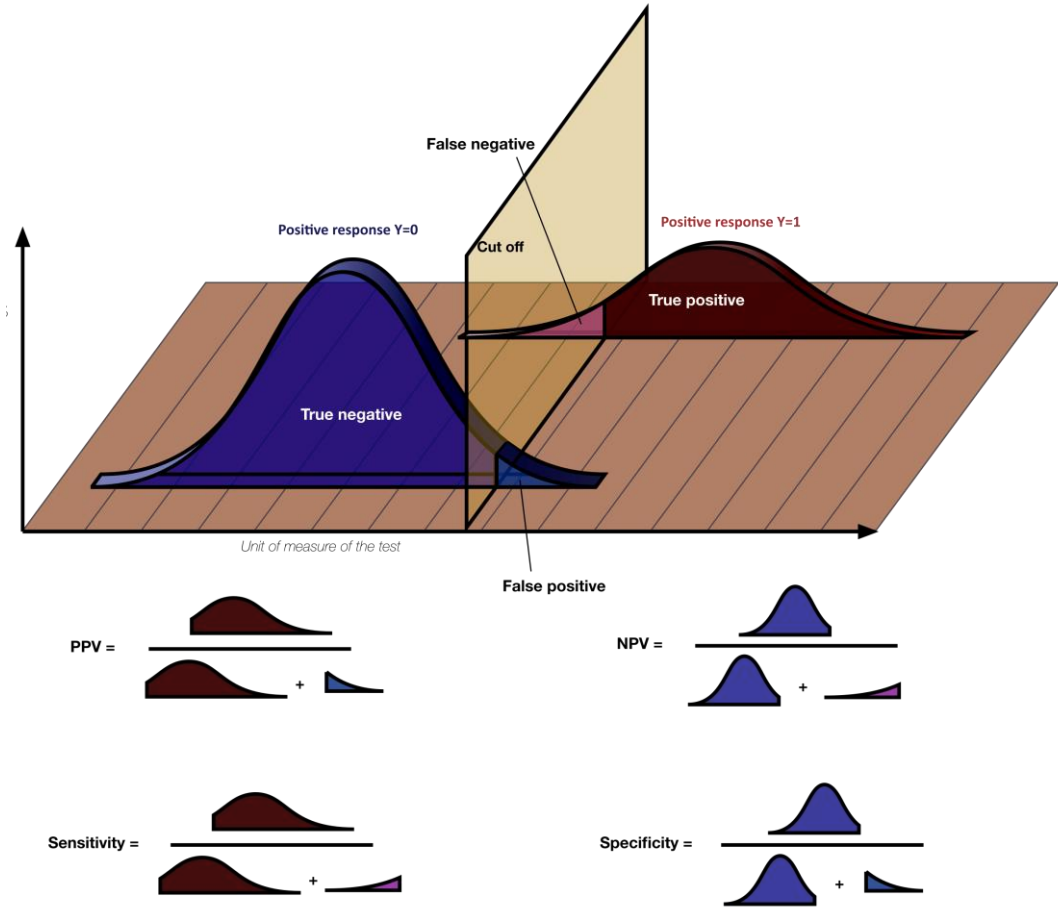
Youden index for a cutoff point x:

$$J(x) = Se(x) + Sp(x) - 1 = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1$$
$$= P(X \leq x|Y = 0) - P(X \leq x|Y = 1)$$

Optimal cutoff: $\boldsymbol{x^* = \max_x J}$

i.e. maximizing the vertical distance between the diagonal and the ROC curve:





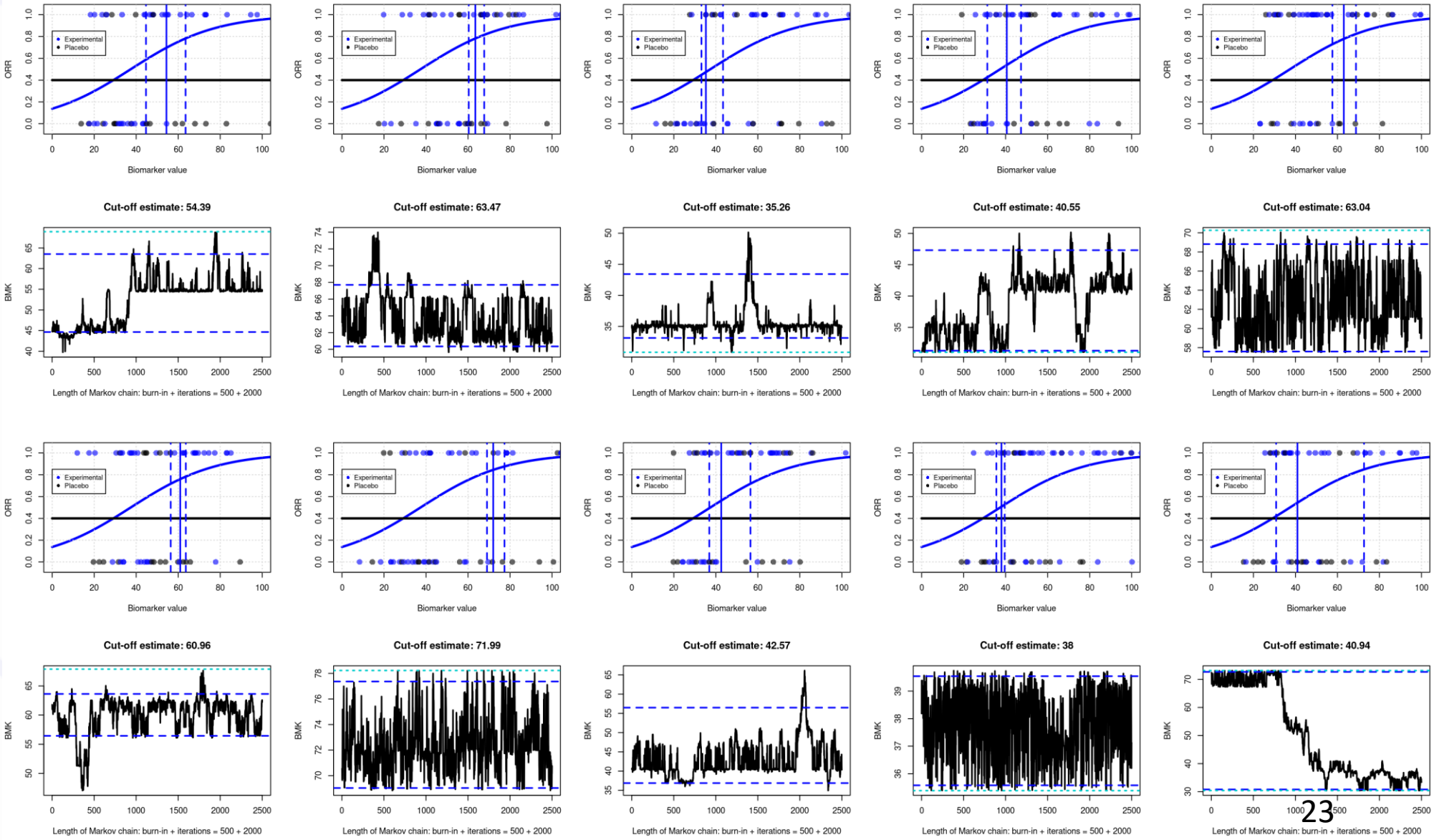
Youden (1950), Cancer https://doi.org/10.1002/1097-0142 & Fluss et al. (2005), Biom. J. https://doi.org/10.1002/bimj.200410135

SARYGA
STATISTICS & METHODOLOGY

Method explored: Bayesian Hierarchical Model (BHM)

Data: 40:20, HPNP1_50,

10 different runs of BHM on different datasets

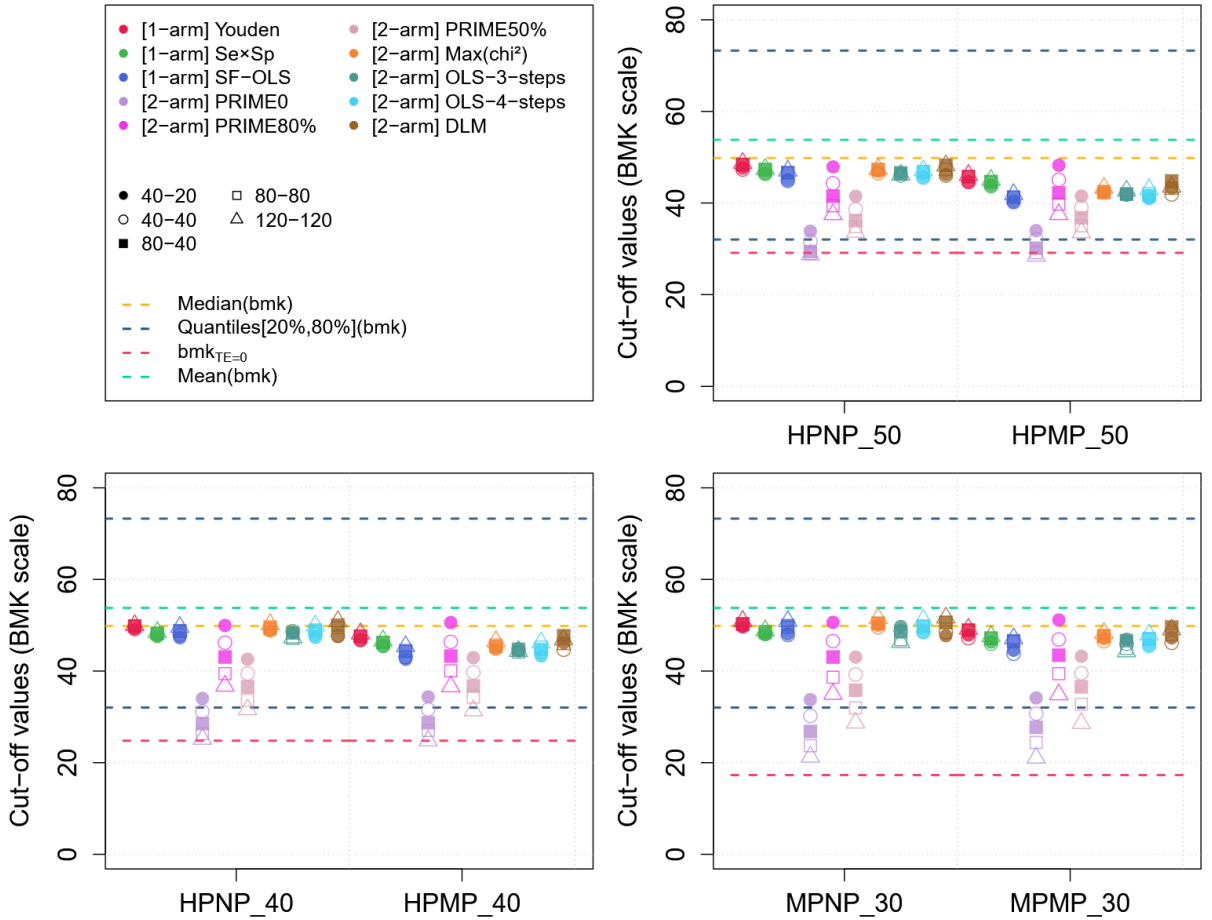


SARYGA
STATISTICS & METHODOLOGY

Sample size sensitivity analysis

	One-arm methods			Two-arm methods						
	Youden	Se×Sp	SF-OLS	PRIME0	PRIME80%	PRIME50%	two-arm-Max(χ^2)	OLS-3-steps	OLS-4-steps	DLM
	Change in $\sqrt{\text{MSE}}$ (compared to previous sample size)									
40-20	10.0	7.7	14.4	15.8	29.5	21.9	14.2	14.1	14.5	14.0
40-40	-0.0	0.1	-0.1	-2.1	-4.3	-3.2	0.8	0.1	0.6	-0.7
80-40	-1.2	-1.6	-1.0	-1.4	-3.6	-2.5	-1.1	-1.3	-1.2	1.4
80-80	0.0	0.0	0.1	-1.5	-3.6	-2.6	0.6	-0.1	0.5	-0.9
120-120	-0.7	-0.7	-0.7	-1.2	-3.2	-2.2	-0.9	-1.1	-0.9	-0.0
	Change in bias (compared to previous sample size)									
40-20	-2.0	-1.6	-4.8	10.3	26.0	18.7	-2.9	0.3	-4.4	-8.9
40-40	-0.0	0.0	-0.0	-2.8	-3.9	-3.2	-0.2	0.7	0.2	1.8
80-40	0.9	0.7	1.7	-2.7	-3.1	-2.8	0.7	-0.9	1.0	0.3
80-80	-0.0	0.0	-0.2	-2.0	-3.3	-2.7	-0.1	0.1	0.1	1.3
120-120	0.4	0.3	0.9	-1.7	-3.1	-2.5	0.7	-0.1	0.9	0.9
	Change in standard deviation (compared to previous sample size)									
40-20	9.7	7.6	13.6	11.5	13.7	11.2	13.9	14.0	13.8	10.8
40-40	-0.0	0.1	-0.1	-0.6	-2.0	-1.0	0.7	0.0	0.7	0.4
80-40	-1.1	-1.5	-0.5	-0.1	-1.8	-0.4	-1.0	-1.2	-1.0	1.8
80-80	0.0	0.0	0.0	-0.8	-1.5	-0.8	0.5	-0.1	0.6	-0.4
120-120	-0.7	-0.7	-0.6	-0.7	-1.0	-0.6	-0.7	-1.0	-0.7	0.3
	Change in $P(c > \text{bmk}_{\text{TE}=0})$ (compared to previous sample size)									
40-20	0.99	1.00	0.93	0.79	0.99	0.95	0.96	0.94	0.94	0.98
40-40	-0.00	-0.00	0.00	-0.06	-0.01	-0.02	-0.01	0.00	-0.00	-0.00
80-40	0.01	0.00	0.02	-0.07	-0.01	-0.04	0.01	0.01	0.02	-0.00
80-80	0.00	-0.00	-0.00	-0.05	-0.02	-0.03	-0.00	-0.00	-0.00	-0.00
120-120	0.00	0.00	0.02	-0.05	-0.02	-0.05	0.01	0.01	0.02	0.00
	Change in mean BMK+ population proportion (compared to previous sample size)									
40-20	0.54	0.55	0.58	0.75	0.53	0.63	0.56	0.56	0.58	0.56
40-40	0.00	0.00	0.00	0.04	0.05	0.05	0.00	0.01	-0.00	0.01
80-40	-0.01	-0.01	-0.02	0.03	0.04	0.04	-0.01	-0.00	-0.01	-0.04
80-80	0.00	0.00	0.00	0.03	0.05	0.04	0.00	0.01	-0.00	0.02
120-120	-0.01	-0.00	-0.01	0.02	0.05	0.04	-0.01	0.00	-0.01	-0.01
	Change in $P(c >$ Analytical cut-off$)$ (compared to previous sample size)									
40-20	0.40	0.40	0.36	0.79	0.99	0.95	0.40	0.49	0.36	0.21
40-40	0.00	0.00	0.00	-0.06	-0.01	-0.02	-0.01	0.01	0.00	0.05
80-40	0.03	0.03	0.04	-0.07	-0.01	-0.04	0.02	-0.02	0.03	0.05
80-80	-0.00	0.00	-0.00	-0.05	-0.02	-0.03	-0.00	0.01	0.00	0.02
120-120	0.02	0.02	0.02	-0.05	-0.02	-0.05	0.02	0.00	0.02	0.03

Operating characteristics for the sample size sensitivity analysis, with differing sample size: 40:20, 40:40, 80:40, 80:80 and 120:120. For 40:20, the values provided are the average metric across the six simulation scenarios
For 40:40, 80:40, 80:80 and 120:120, the values provided are the difference with the previous (i.e. smaller) sample size setting.



SARYGA
STATISTICS & METHODOLOGY

Biomarker distribution sensitivity analysis

	One-arm methods			Two-arm methods						
	Youden	Se×Sp	SF-OLS	PRIME0	PRIME80%	PRIME50%	two-arm-Max(χ^2)	OLS-3-steps	OLS-4-steps	DLM
	$\sqrt{\text{MSE}}$									
Skewed Gamma	9.7	7.4	14.2	12.3	30.9	19.5	14.9	14.8	14.5	14.8
Medium-skewed Gamma	9.7	7.5	14.1	13.9	28.4	20.0	14.2	13.9	14.2	14.1
Less-skewed Gamma	10.0	7.7	14.4	15.8	29.5	21.9	14.2	14.1	14.5	14.0
Uniform	12.8	11.0	16.4	16.1	28.6	21.8	16.5	16.7	16.7	16.0
	Bias									
Skewed Gamma	-1.1	-0.0	-4.0	3.6	22.4	12.6	-2.8	-3.1	-4.3	-10.1
Medium-skewed Gamma	-1.8	-1.0	-4.6	7.3	23.5	15.7	-3.3	-1.4	-4.5	-9.2
Less-skewed Gamma	-2.0	-1.6	-4.8	10.3	26.0	18.7	-2.9	0.3	-4.4	-8.9
Uniform	-2.3	-2.5	-3.2	7.7	25.2	17.4	-0.9	-0.7	-2.7	-6.2
	Standard deviation									
Skewed Gamma	9.6	7.4	13.7	11.6	21.2	14.7	14.6	14.5	13.9	10.8
Medium-skewed Gamma	9.5	7.4	13.3	11.5	15.9	12.1	13.8	13.8	13.5	10.7
Less-skewed Gamma	9.7	7.6	13.6	11.5	13.7	11.2	13.9	14.0	13.8	10.8
Uniform	12.6	10.8	16.0	13.8	13.5	12.9	16.4	16.6	16.5	14.7
	$P(c > \text{bmk}_{\text{TE}=0})$									
Skewed Gamma	0.70	0.68	0.72	0.62	0.94	0.83	0.74	0.76	0.71	0.64
Medium-skewed Gamma	0.95	0.96	0.86	0.73	0.98	0.92	0.88	0.88	0.86	0.90
Less-skewed Gamma	0.99	1.00	0.93	0.79	0.99	0.95	0.96	0.94	0.94	0.98
Uniform	0.93	0.95	0.81	0.71	0.98	0.91	0.85	0.84	0.81	0.88
	Mean BMK+ population proportion									
Skewed Gamma	0.45	0.47	0.41	0.51	0.35	0.38	0.40	0.39	0.42	0.47
Medium-skewed Gamma	0.50	0.52	0.52	0.66	0.45	0.53	0.50	0.49	0.52	0.52
Less-skewed Gamma	0.54	0.55	0.58	0.75	0.53	0.63	0.56	0.56	0.58	0.56
Uniform	0.55	0.56	0.59	0.69	0.53	0.59	0.57	0.57	0.59	0.57
	$P(c >$ Analytical cut-off$)$									
Skewed Gamma	0.43	0.47	0.38	0.62	0.94	0.83	0.41	0.41	0.37	0.19
Medium-skewed Gamma	0.40	0.43	0.36	0.73	0.98	0.92	0.39	0.44	0.36	0.21
Less-skewed Gamma	0.40	0.40	0.36	0.79	0.99	0.95	0.40	0.49	0.36	0.21
Uniform	0.42	0.41	0.42	0.71	0.98	0.91	0.47	0.47	0.43	0.34

Operating characteristics for the biomarker distribution sensitivity analysis, with differing distributions: uniform, "skewed Gamma", "medium-skewed Gamma", "less-skewed Gamma". The values provided are the average metric across the six simulation scenarios with a 40:20 sample size.

SARYGA
STATISTICS & METHODOLOGY