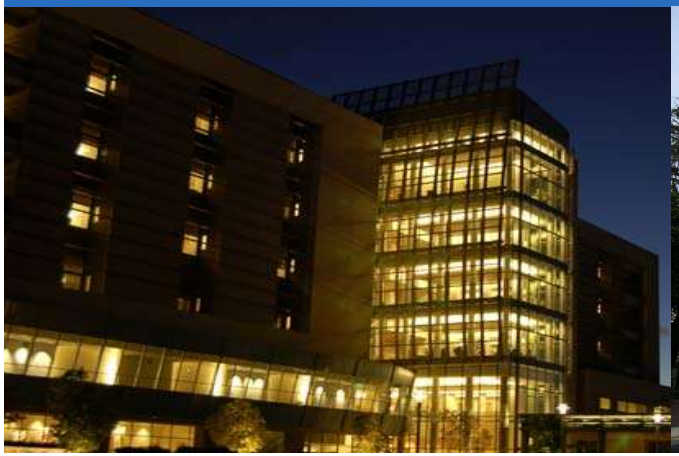
Integrating Biomarkers into Targeted Therapy and Immunotherapy Clinical Trials

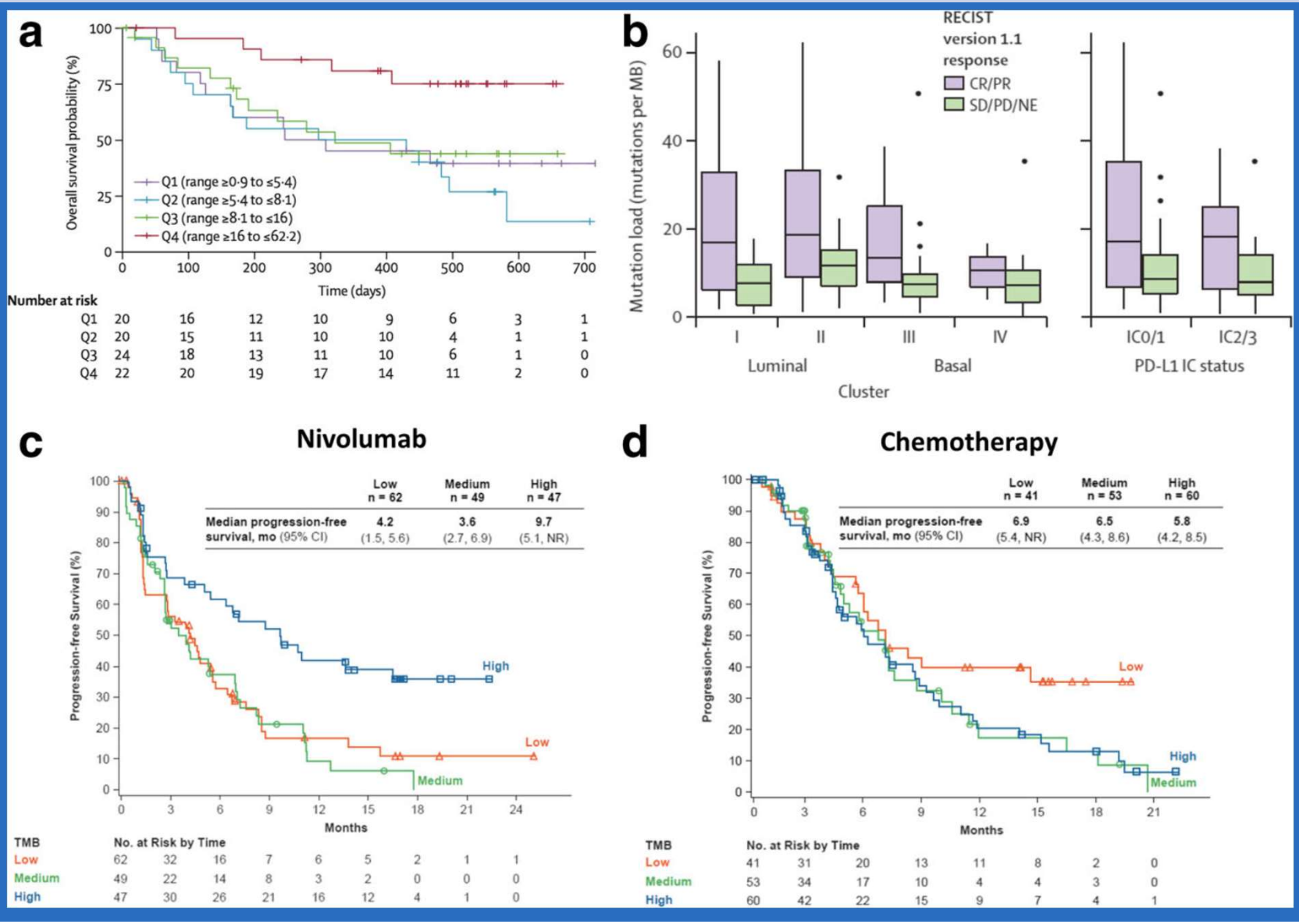
Paul Frankel, PhD
Research Professor, Biostatistics
City of Hope Comprehensive Cancer Center



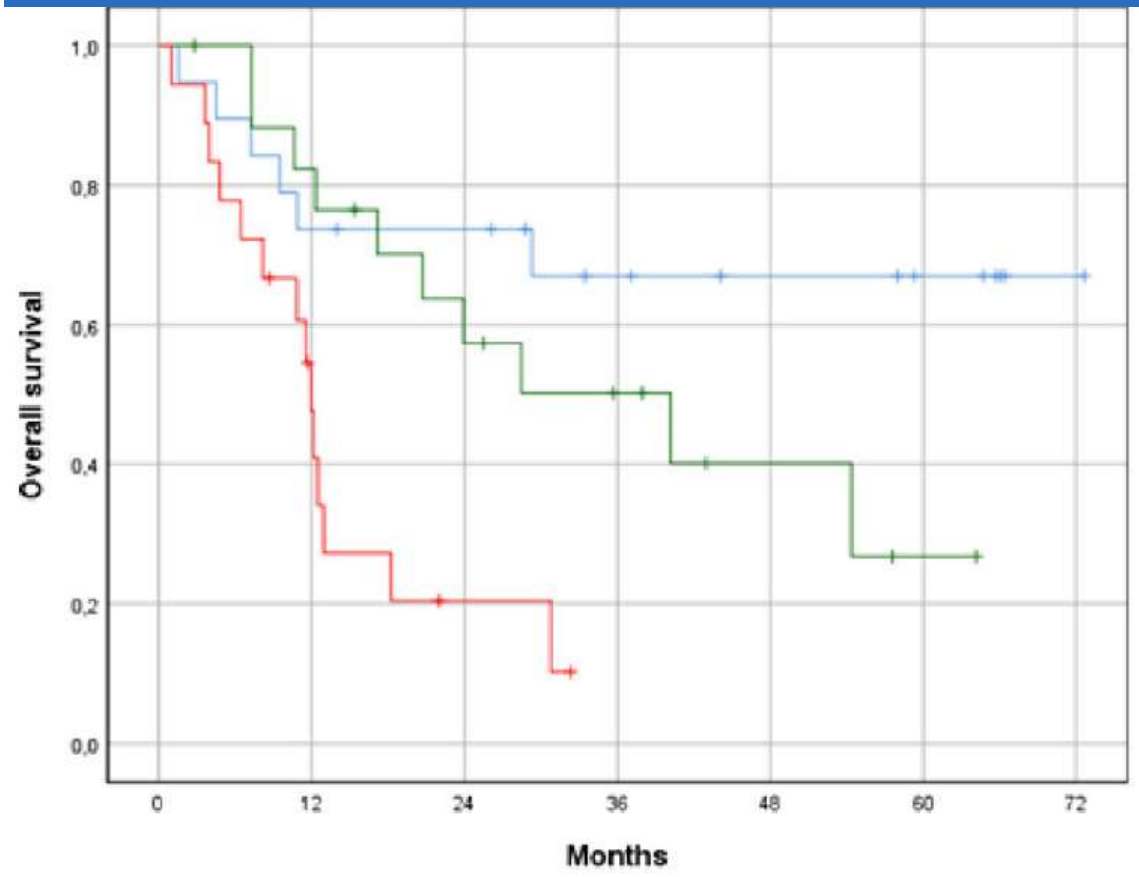
Integrating Biomarkers

- Examples
- Prognostic v Predictive
- Mirvetuximab
  - Example of Problems with Pooling
  - Example of Other Major factors
- Decision Making (S vs S+E)
- Decision Making (S vs E)

Immunotherapy in Bladder (Aggen, 2017)



PET Biomarkers in Anti-PD1 for Melanoma (Seban, 2019)



Low-risk group

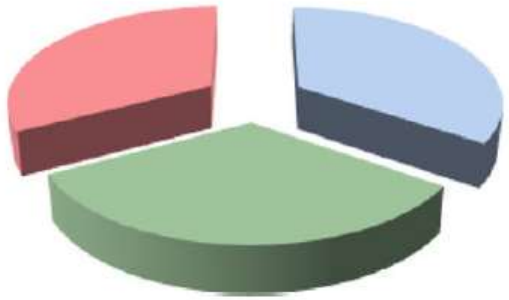
Low TMTV and low BLR, n=19 pts

Intermediate-risk group

High TMTV or high BLR, n=18 pts

High-risk group

High TMTV and high BLR, n=18 pts



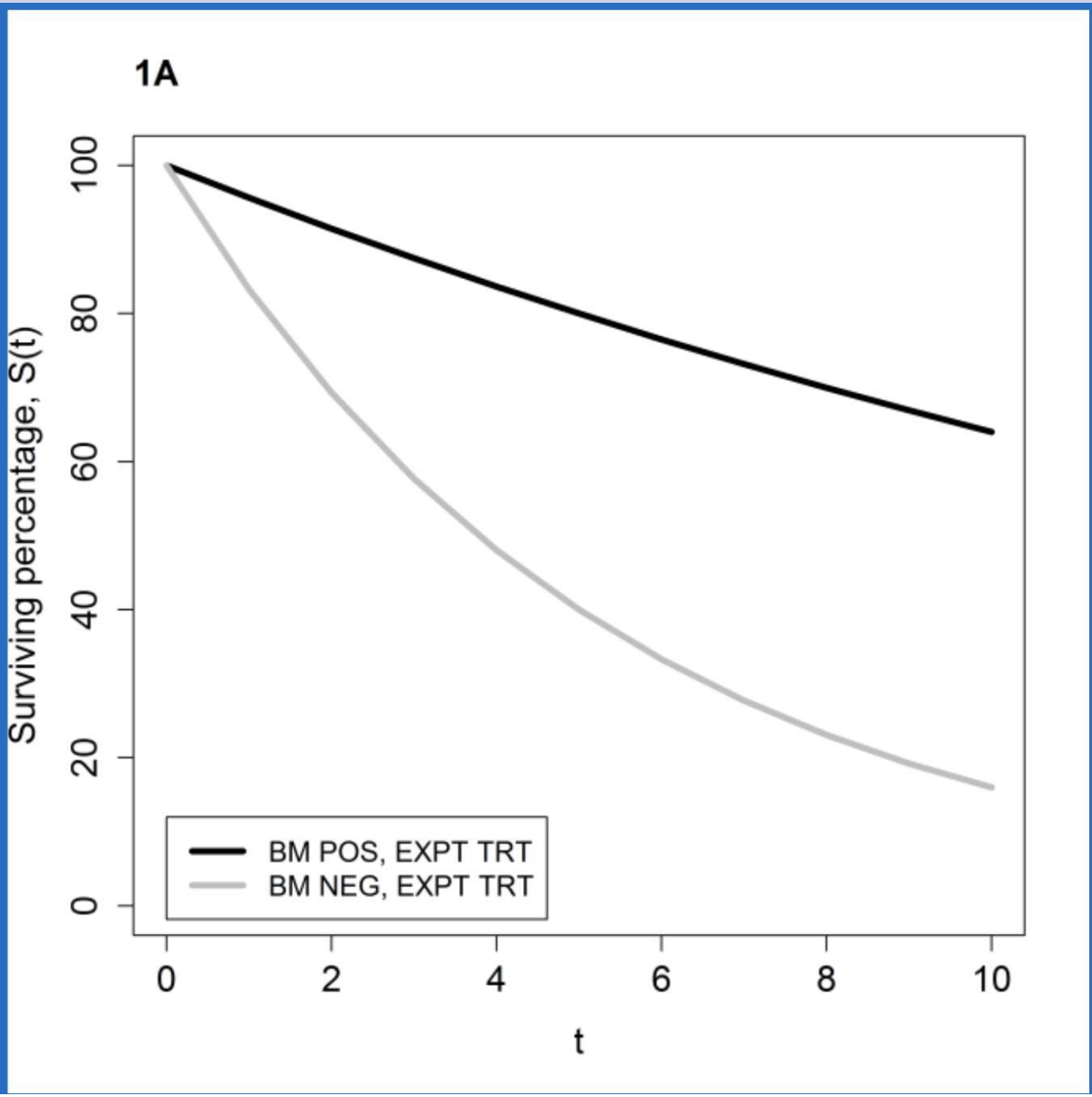
Integrating Biomarkers

- Examples√
- Prognostic v Predictive
- Mirvetuximab
  - Example of Problems with Pooling
  - Example of Other Major factors
- Decision Making (S vs S+E)
- Decision Making (S vs E)

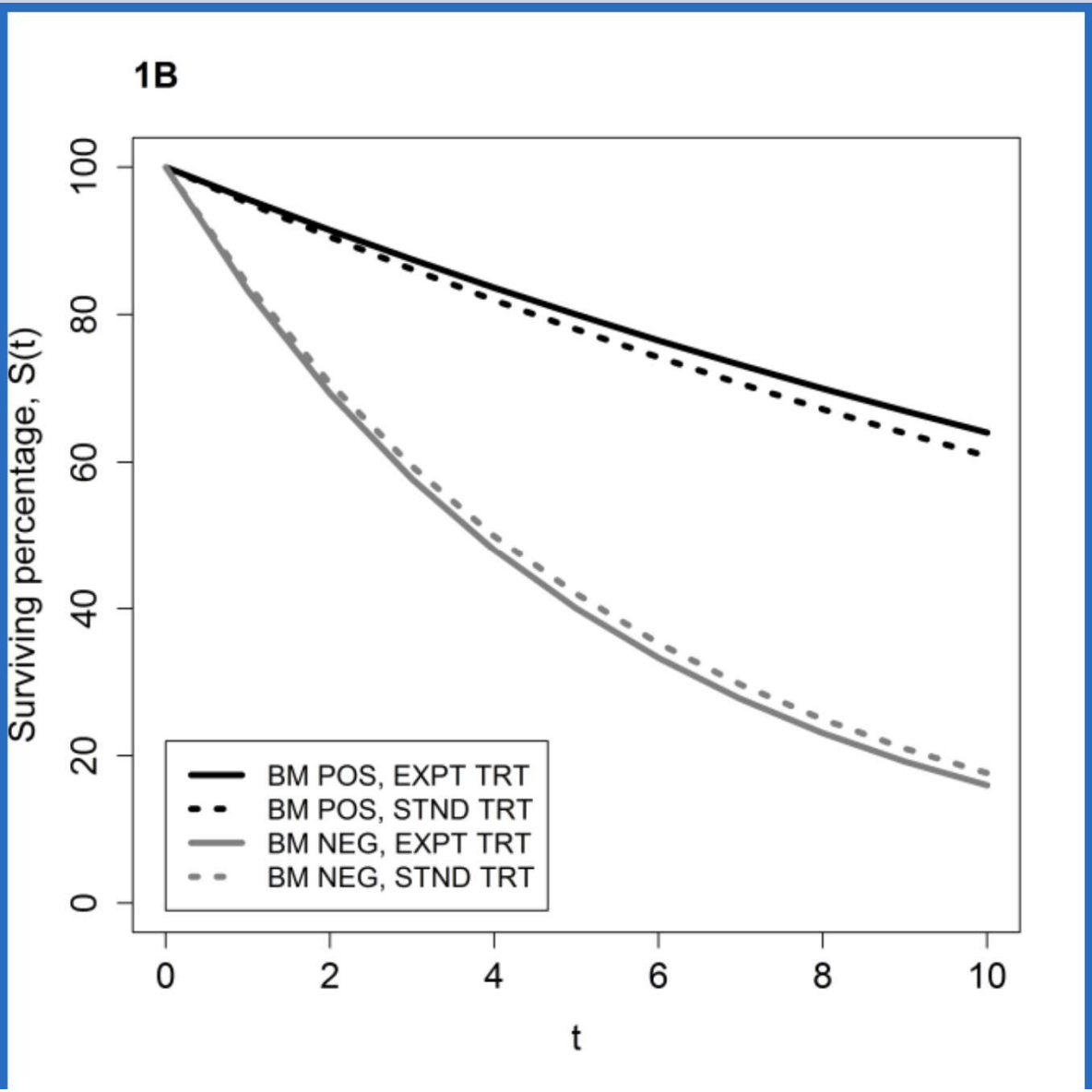
Prognostic vs Predictive

- Prognostic: Provides information about patients outcome, regardless of therapy
- Predictive: Provides information about the effect of a therapeutic intervention

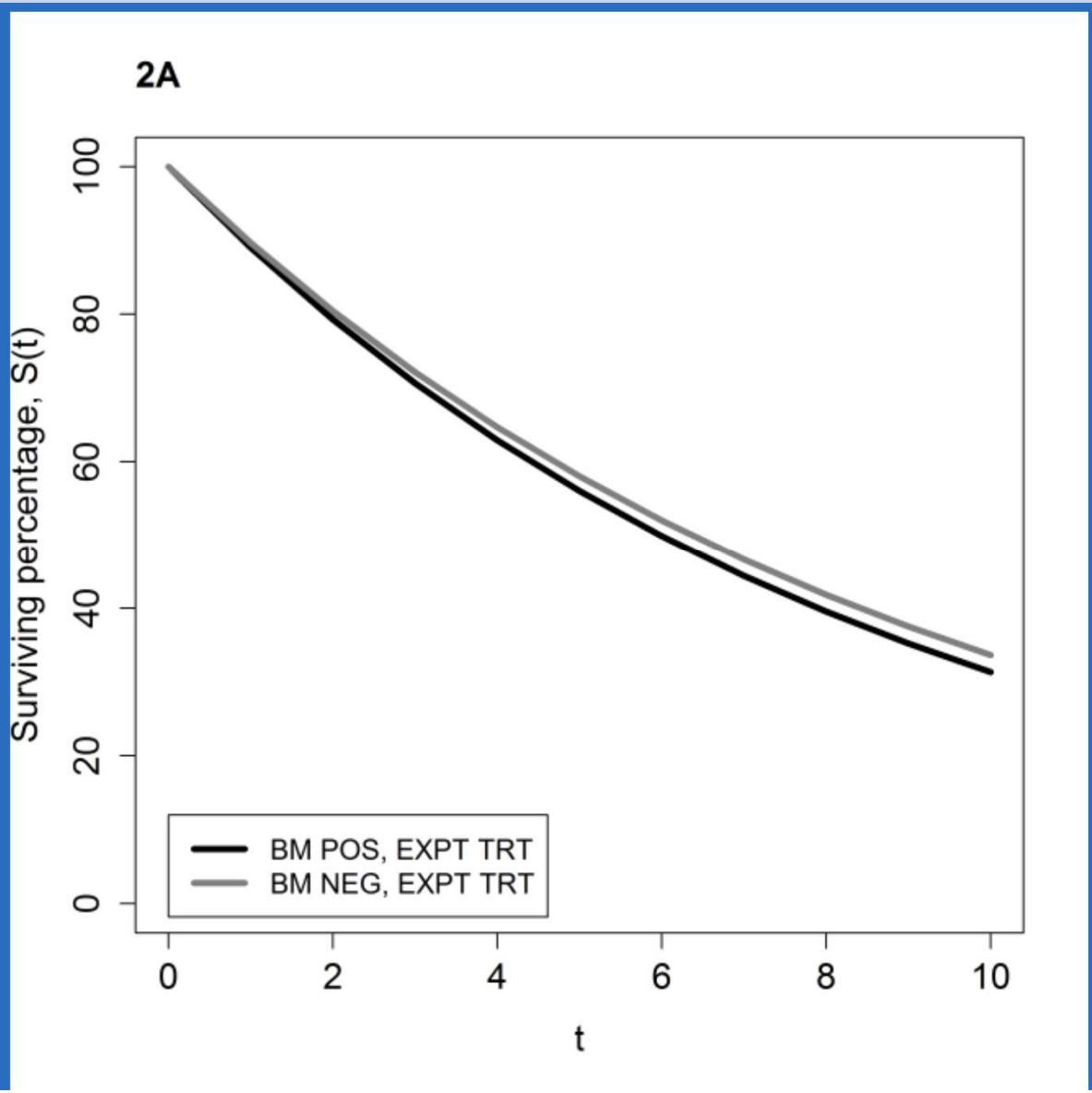
Prognostic but Not Predictive (A)



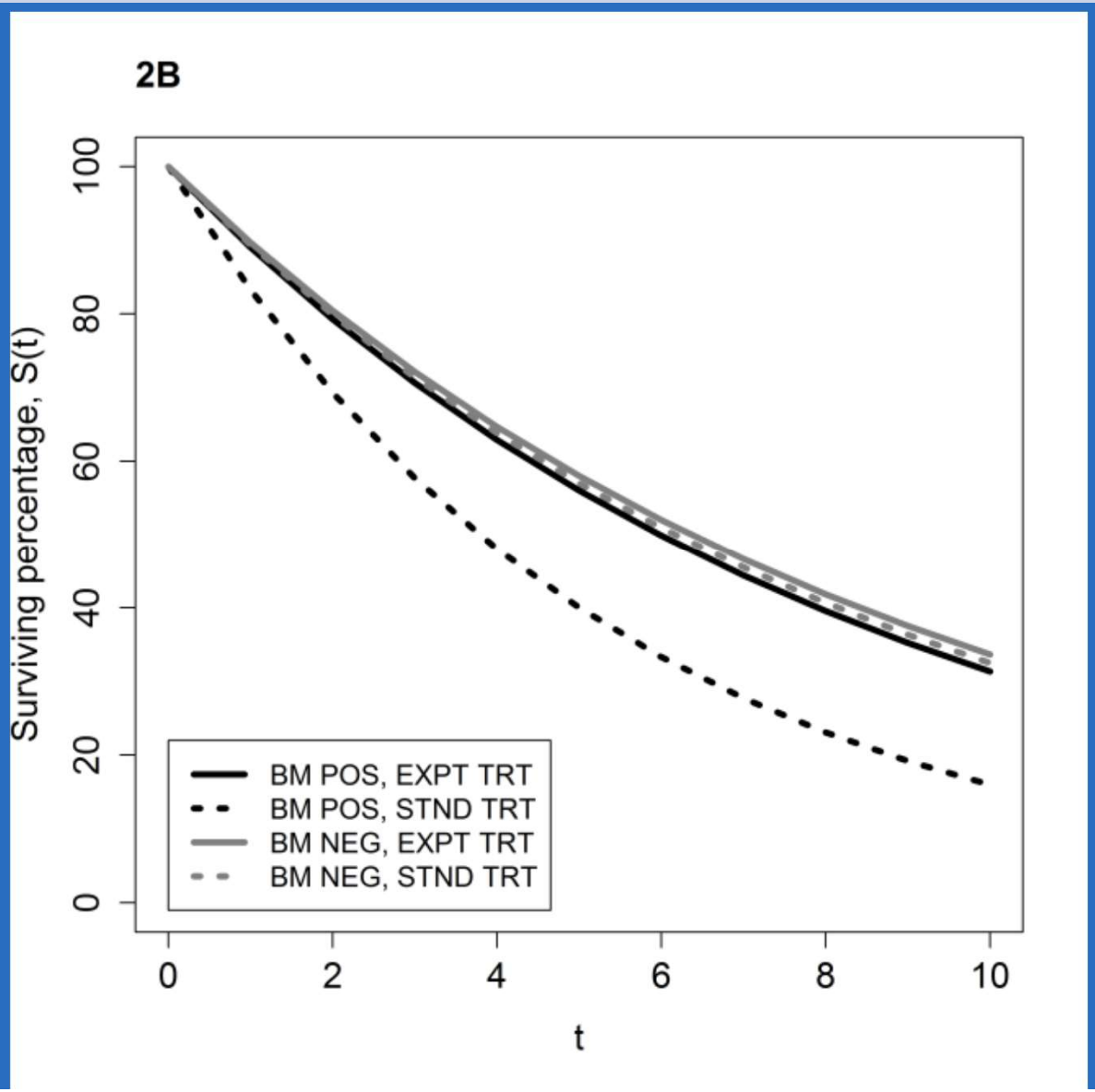
Prognostic but Not Predictive (B)



Prognostic (neg) AND Predictive (A)



Prognostic (neg) AND Predictive (B)



Integrating Biomarkers

- Examples√
- Prognostic v Predictive√
- Mirvetuximab
  - Example of Problems with Pooling
  - Example of Other Major factors
- Decision Making (S vs S+E)
- Decision Making (S vs E)

Mirvetuximab – Phase I

Table 3
Relationship of antitumor activity with archival FRα expression level.

FRα expression	No. of patients	CR	PR	ORR N (%)	PFS (months)	PFS 95% CI
Low	6	0	0	0 (0.0)	2.8	(1.3, 5.4)
Medium	5	0	1	1 (20.0)	3.9	(2.6, 12.7)
High	16	2	3	5 (31.3)	5.4	(2.8, −)
Overall	27	2	4	6 (22.2)	4.2	(2.8, 5.4)

CR, complete response; PR, partial response; ORR, objective response rate; PFS, progression-free survival; CI, confidence interval.

Mirvetuximab – Forward I

N=336 (2:1), High N=218, Med N=118

	Mirvetuximab	chemo	Hazard Ratio	'P'
Response rate	22%	12%	--	0.15
PFS	--	--	0.98	0.9
High Group RR	24%	10%	--	0.014
High Group PFS			0.69	0.049
High Group OS			0.62	0.033

We can infer that if the overall HR=0.98, and the HR=0.69 for "High", which was the largest group, the HR had to be very poor for the medium group.

Secondary analysis in High group did not meet 0.025 significance needed for subgroup analysis (unpublished)

Mirvetuximab – Pooled Analysis

Apparently, there was no difference in the pooled analysis across medium/high groups (unpublished, ASCO Abstract)

What was their mistake?

- **Main Cause:**

  Prediction is very difficult. Especially about the future.

- **…. Biomarker issues are very challenging**
- **…. Biomarker issues are made more complex when pooling studies.**
- **…. Biomarkers are more challenging in a complex clinical situation where other factors can play a large role**

Example of Cross-Study Differences

Phase II study of carboplatin and paclitaxel with or without vorinostat in first line advanced non-small cell lung cancer

Ramalingam SS, Maitland M, Frankel P, Argiris A, Koczywas M, Gitlitz B, Espinoza-Delgado I, Vokes EE, Gandara DR, Belani CP.

California Cancer Consortium

University of Chicago Cancer Consortium

Emory University

City of Hope

First: Study

Phase II by NCI-CTEP (N=62 vs N=32)

	Vorinostat +chemo	Placebo +chemo	Hazard Ratio	'P'
Response rate (RECIST)	**34%**	**12.5%**		**0.021**
Progression-Free Survival	**6.0 m (median)**	**4.1 m (median)**	**0.79**	**0.33**
Overall Survival	**13 m (median)**	**9.7 m (median)**	**0.67**	**0.17**
1-year survival rate	**53%**	**35%**		

Second Study

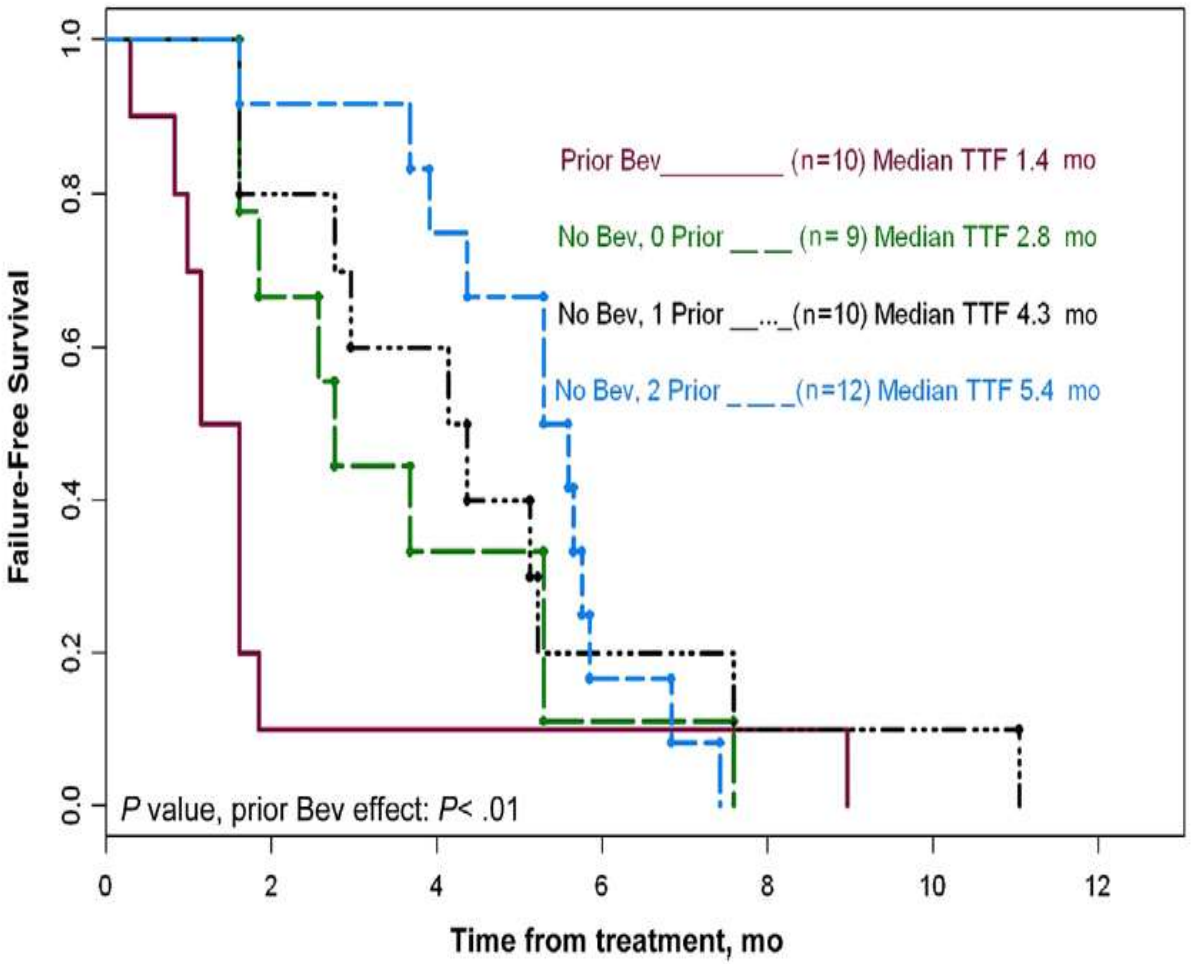
Phase II/III Study by Merck/CRO (N=126 vs 127)*

	Vorinostat +chemo	Placebo +chemo	'P'
Response rate (RECIST)	**28/125 22%**	**36/123 29%**	**0.14**
Progression-Free Survival	**4.3 m (median)**	**5.5 m (median)**	**0.86**
Overall Survival	**11 m (median)**	**14 m (median)**	**0.99**

*Unpublished, but reported to clinicaltrials.gov. Extent of variation between studies can be very surprising. Opportunity for unanticipated problems for evaluating biomarkers in pooled studies.

Vinorelbine/Sorafenib BC (Luu, 2014) – Other Factors

Figure 2 Time to Treatment Failure Based on Prior Bevacizumab Treatment and Prior Chemotherapy



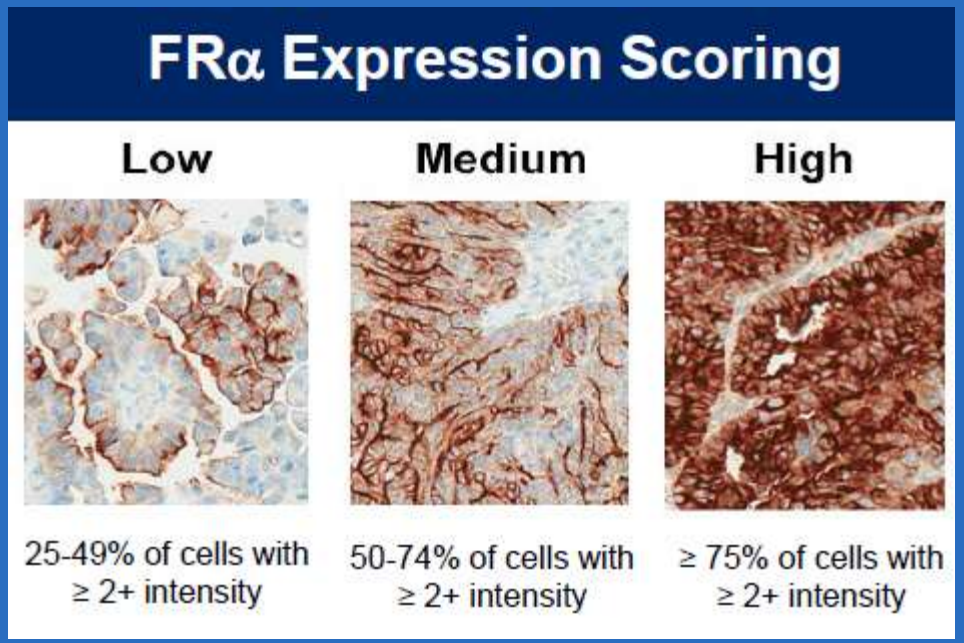
What if high group had different prior therapy than low group?

Integrating Biomarkers

- Examples√
- Prognostic v Predictive√
- Mirvetuximab√
  - Example of Problems with Pooling√
  - Example of Other Major factors√
- Decision Making (S vs S+E)
- Decision Making (S vs E)

Decision Making

- What determines the cut-point for biomarker positive or biomarker negative? IMGN853 chose ≥50% with 2+ intensity:



Decision Making (S vs S+E)

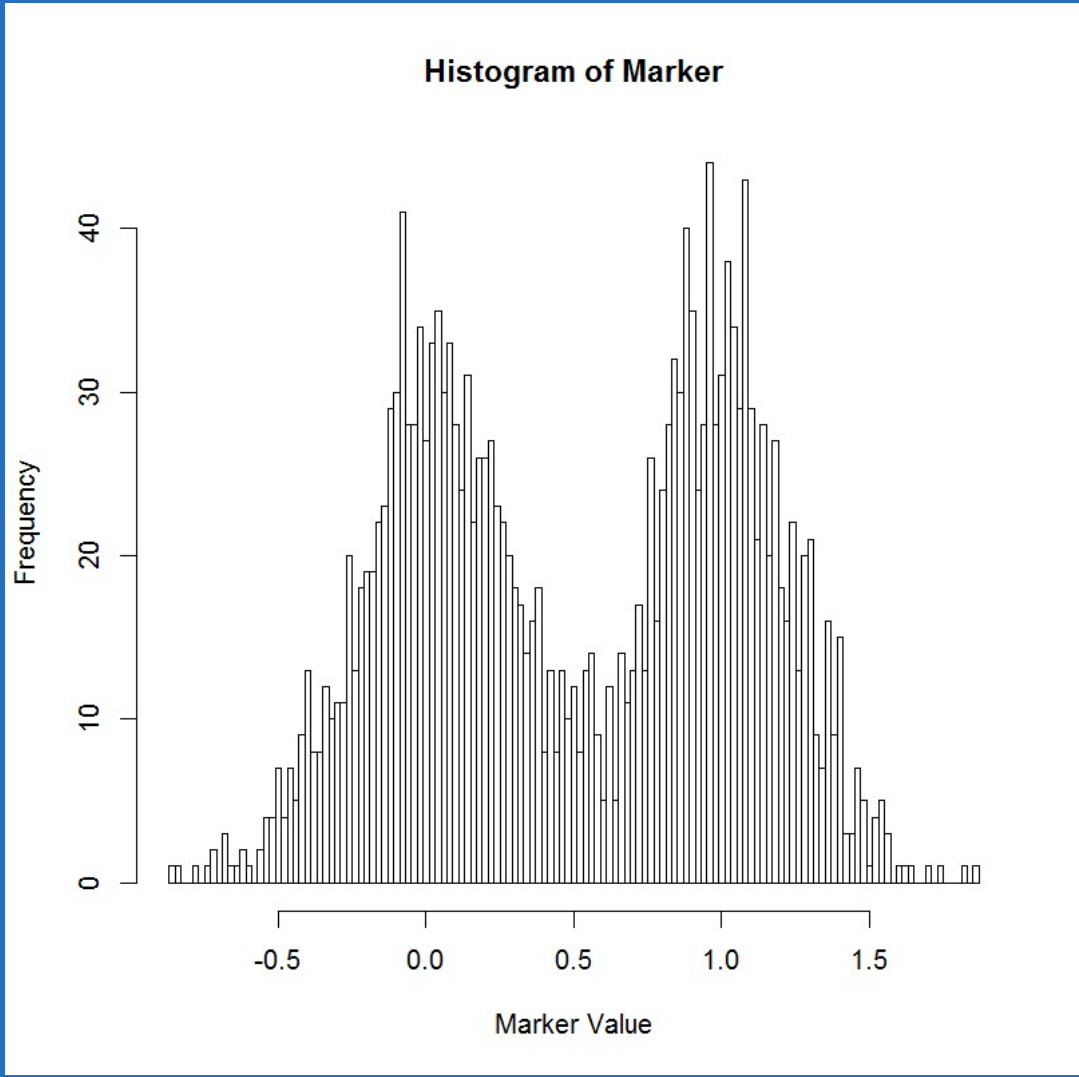
Simulation:

There are two groups: Positive and Negative, with positive benefitting from targeted therapy.

BUT there is measurement error. With scaling (e.g. subtracting a standard) positives score a mean of 1 (SD 0.25), negatives score 0 (SD 0.25).

Assume Neg has no benefit from targeted therapy, Pos has HR=0.66 (exp)

Standard+Exp vs Standard

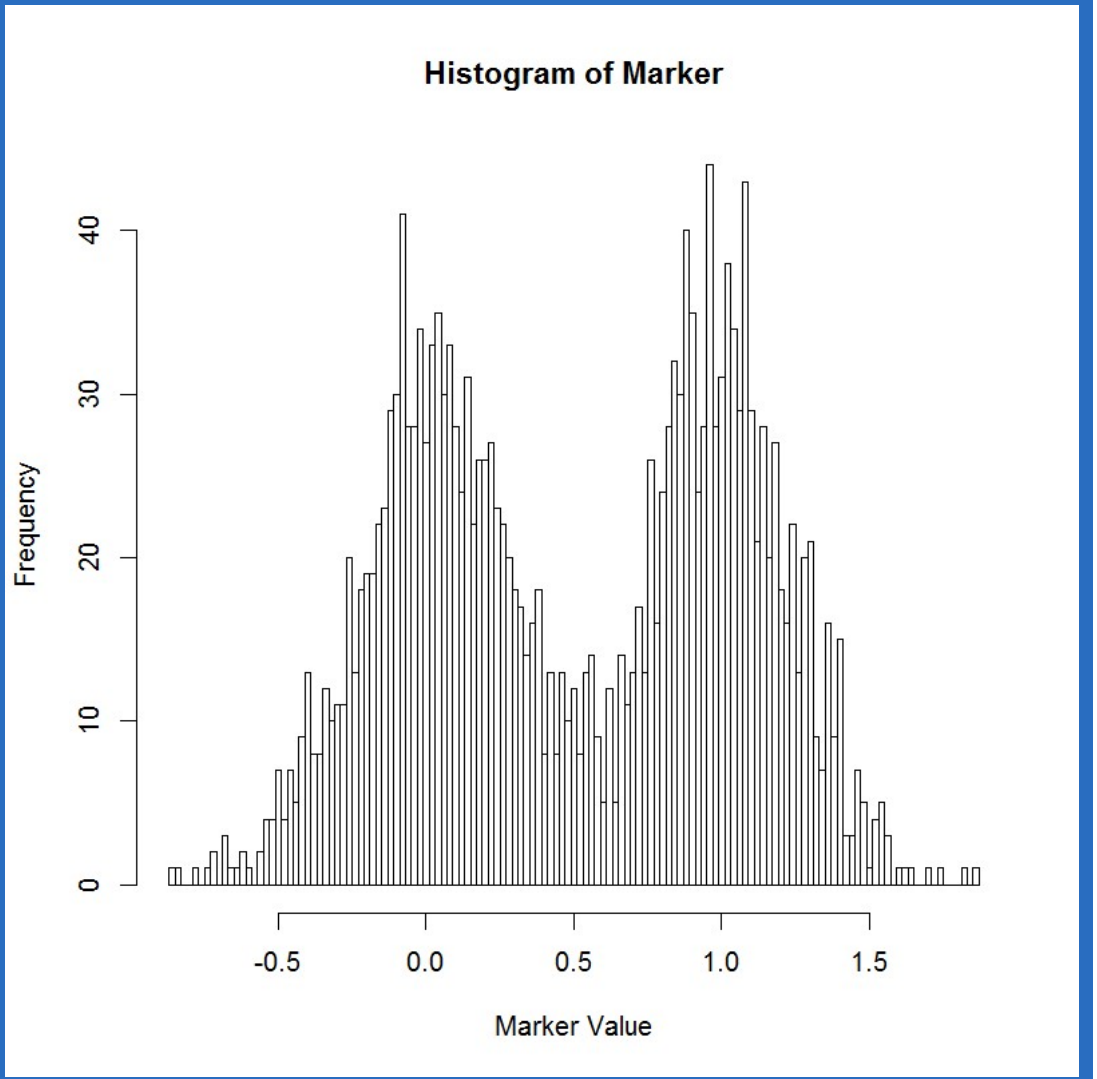


Decision Making S vs S+E

Simulation:

What is the best cut-point?

1) 0.5? Best power?
2) Less than 0.5? Captures some additional positive patients, at the cost of diluting signal and possibly a little more toxicity?
3) Lowest value where we still get a positive signal?
4) How well would we estimate cut-point for the best power?

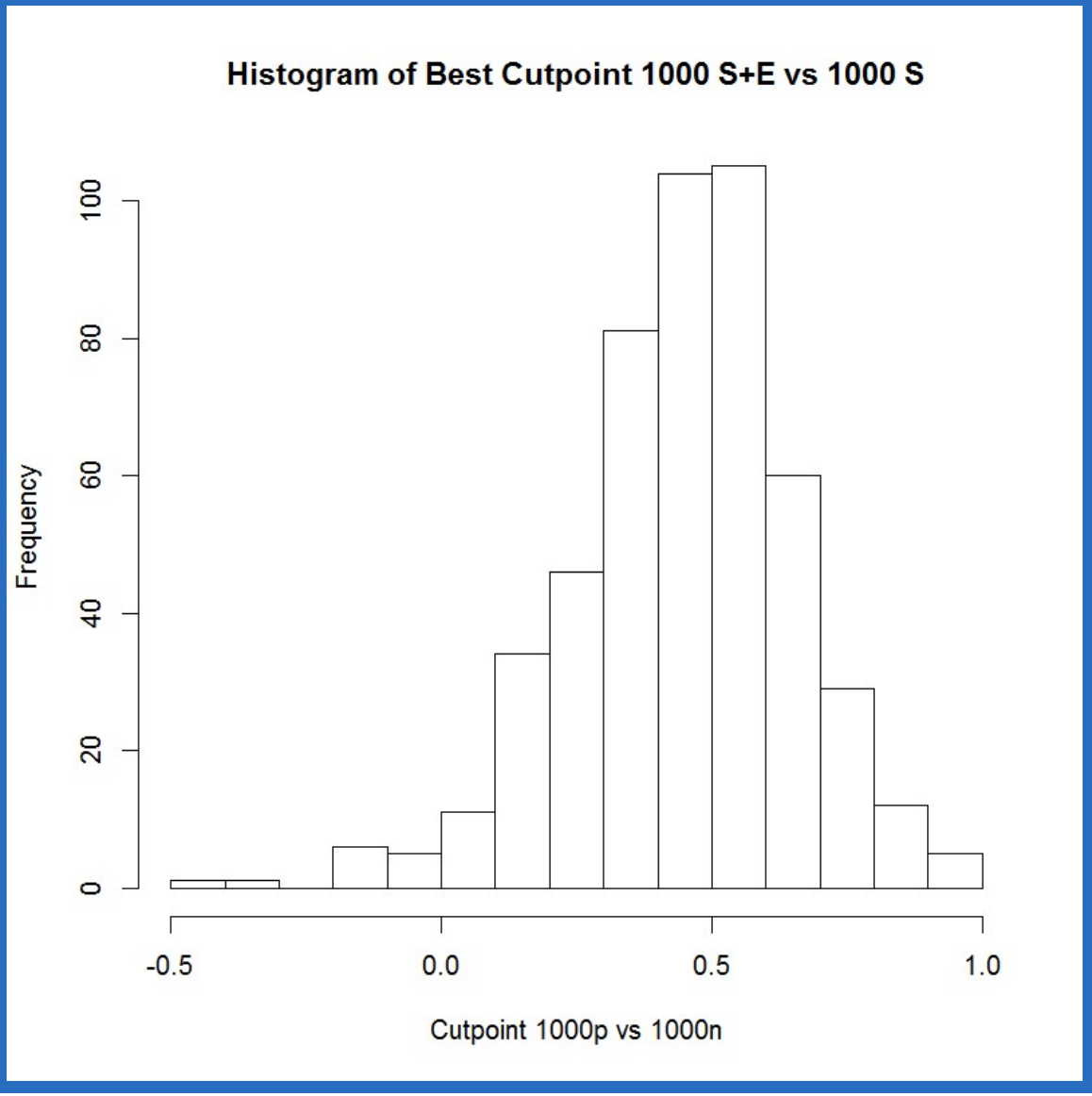


Decision Making (S v S+E)

With 1000 positive patients, 1000 negative patients, each with half treated with S+E,half with S, we get a distribution of best cut-points (500 simulations)

Even with 2000 patients, in this scenario, there is considerable variability in the choice of cutpoint.

This also assumed a nice bimodal distribution. Mean is 0.45.



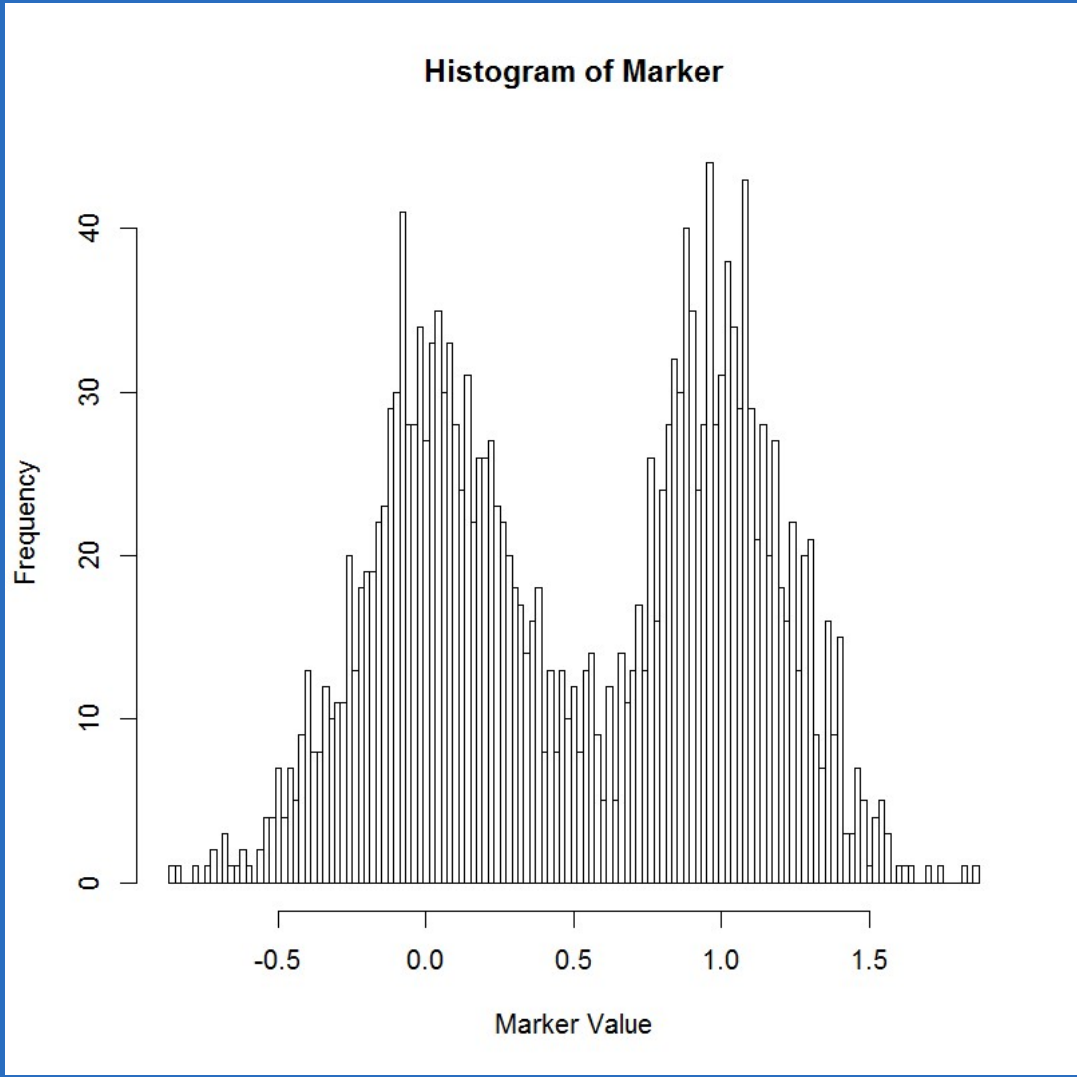
Decision Making S vs E

Simulation:

What is the best cut-point?

Assume 0.66 HR for Pos, but actually S is better for Neg.

1) 0.5? Best power?
2) S vs E is the Mirvetuximab study.



Decision Making S vs E

With 1000 positive patients, 1000 negative patients, each with half treated with S,half with E, we get a distribution of best cut-points (500 simulations)

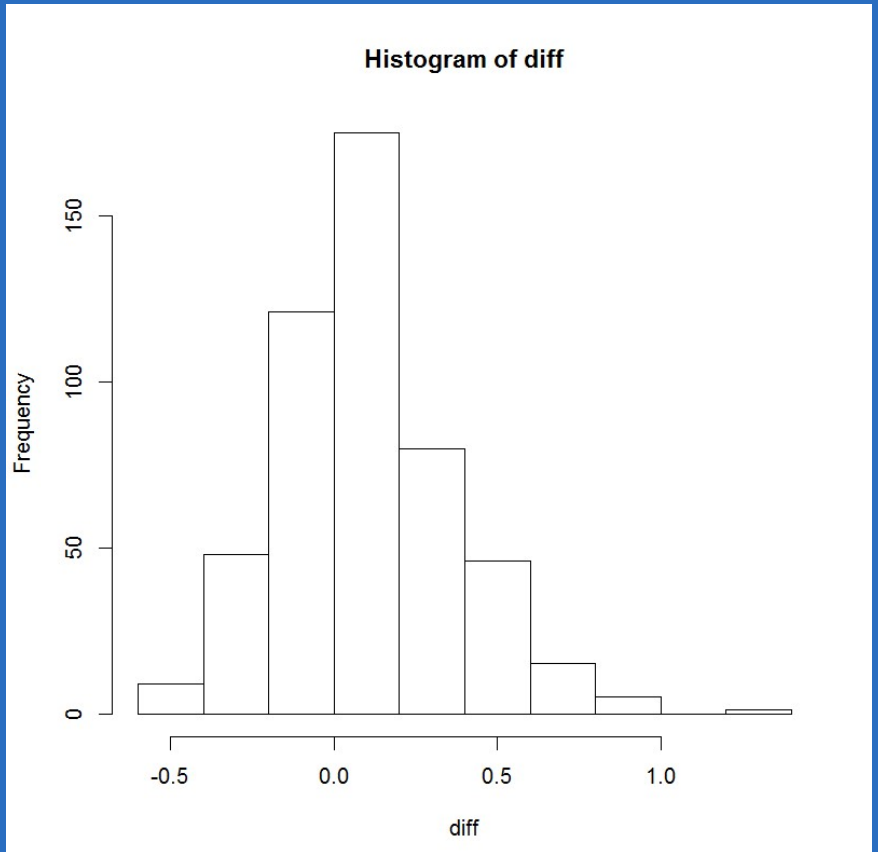
Pos w/ Exp: 250 days (HR=0.66)

Pos w/Standard: 167 days

Neg w/Exp: 143 days

Neg w/Standard: 167 days

The Mean Best Cutpoint goes from 0. 45 to 0.55 as expected from simulation



Integrating Biomarkers

- Examples√
- Prognostic v Predictive√
- Mirvetuximab√
  - Example of Problems with Pooling√
  - Example of Other Major factors√
- Decision Making (S vs S+E) √
- Decision Making (S vs E) √

Conclusion/Questions?

1) Cut-points require detailed discussion supported by simulations to better understand the risks/benefits of different cut-point approaches
2) Due to large confidence intervals in the estimate of the “best” cut-point, we must consider ways to have such cut-points re-evaluated as “real-world” evidence accumulates.