Multiplicity and other issues related to biomarker-based oncology trials
ASA NJ Chapter

Keaven M. Anderson, Christine K. Gause, Cong Chen

Merck Research Laboratories

November 11, 2016

With thanks to Eric Rubin, Lu Wang

MERCK

Abstract

With highly active targeted therapies becoming more commonplace in recent years, strategies for development become increasingly important in terms of time-to-market, clinical trials size and the breadth of patient population that may benefit from a drug. We discuss strategies and corresponding statistical tools that have been applied. Early single-arm trials followed quickly by randomized trials is a bedrock approach for many indications, but not without pitfalls. The possibility of doing trials that are target-based rather than histology-based as well as other novel approaches are considered. Another challenge is developing one or more biomarkers at the same time that a drug is being developed.

MERCK

Overview

- Examples of completed and ongoing biomarker-based studies
  - Generally based on checkpoint inhibitor MK-3475/pembrolizumab/Keytruda
- Progression of studies for a cancer type
- Companion vs. complementary diagnostic device
- Multiplicity control with the graphical method and group sequential design

MERCK

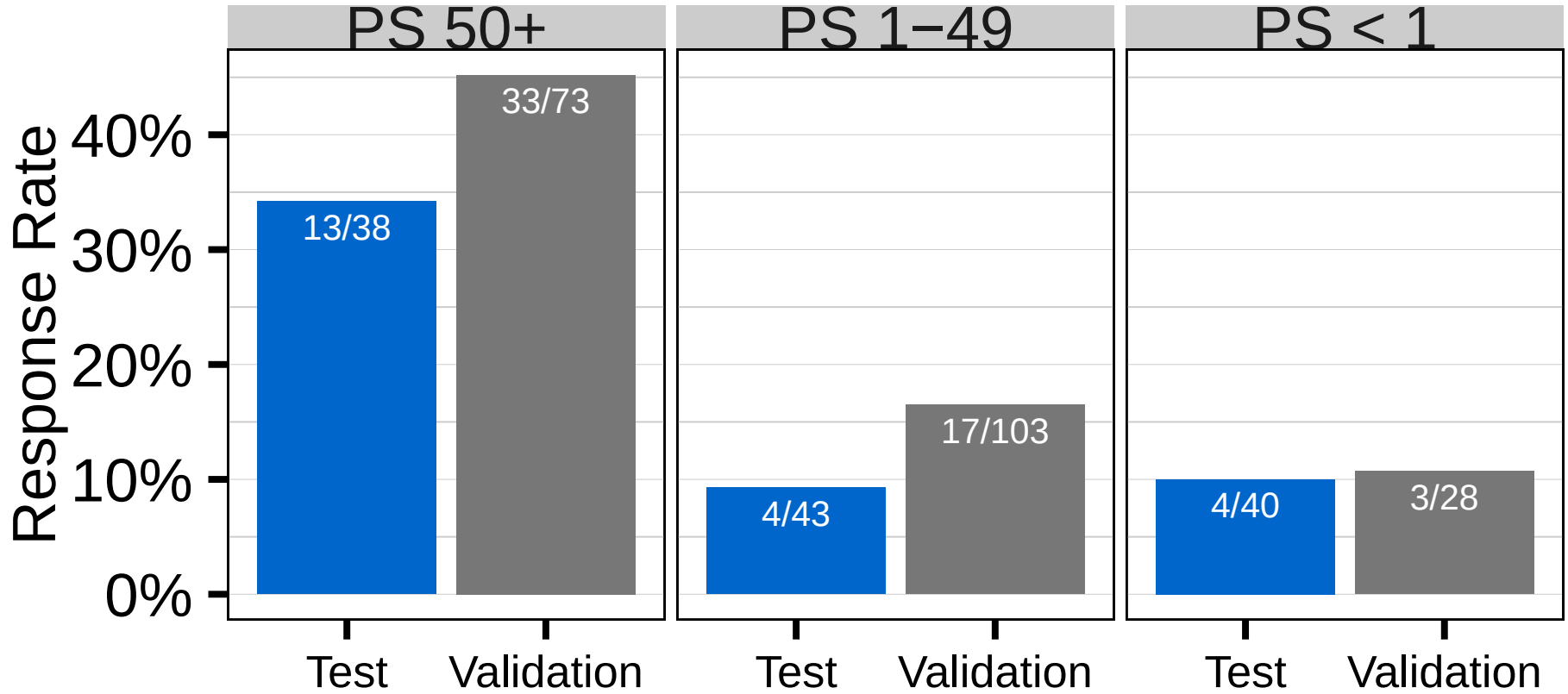
PD-1: Programmed Cell Death Protein 1

- Source: Wikipedia
  - PD-1 is a protein and cell-surface receptor
  - Binds to 2 ligands: PD-L1 and PD-L2
  - "PD-1, functioning as an immune checkpoint, plays an important role in down regulating the immune system by preventing the activation of T-cells, which in turn reduces autoimmunity and promotes self-tolerance."
  - "A new class of drugs that block PD-1, the PD-1 inhibitors, activate the immune system to attack tumors..."
  - "Many tumor cells express PD-L1, an immunosuppressive PD-1 ligand; inhibition of the interaction between PD-1 and PD-L1 can enhance T-cell responses in vitro and mediate preclinical antitumor activity."
- Pembrolizumab and nivolumab are examples of antibodies that inhibit PD-1
- Each has been studied in patients with a diagnostic measuring PD-L1

MERCK

Keynote 001: Objective Response by Biomarker Level

Pembrolizumab for non-small-cell lung cancer; PS=proportion score
Early trial leading to randomized trial design

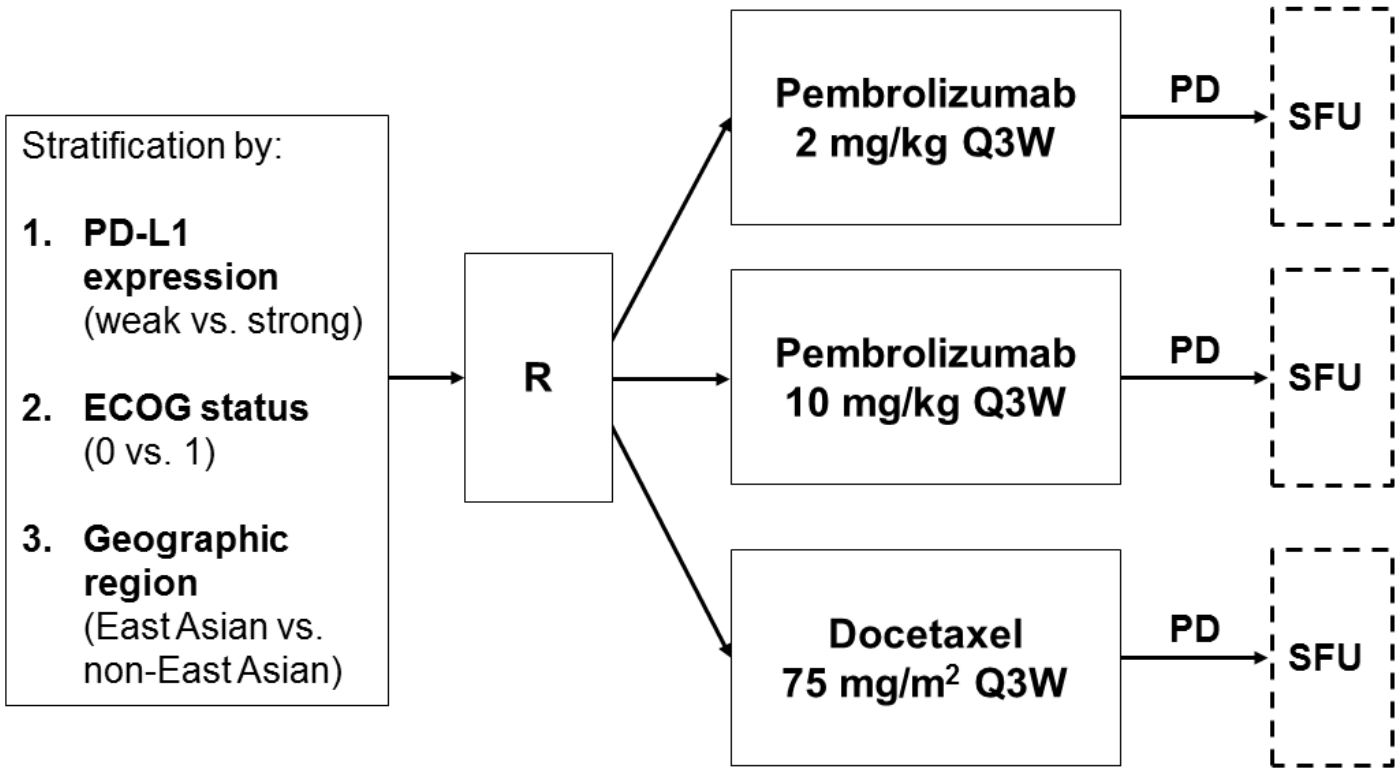


Garon et al. [2015]

MERCK

Keynote 010: Design (NSCLC, biomarker-based)

Previously Treated PD-L1 Positive Advanced Non-Small-Cell Lung Cancer

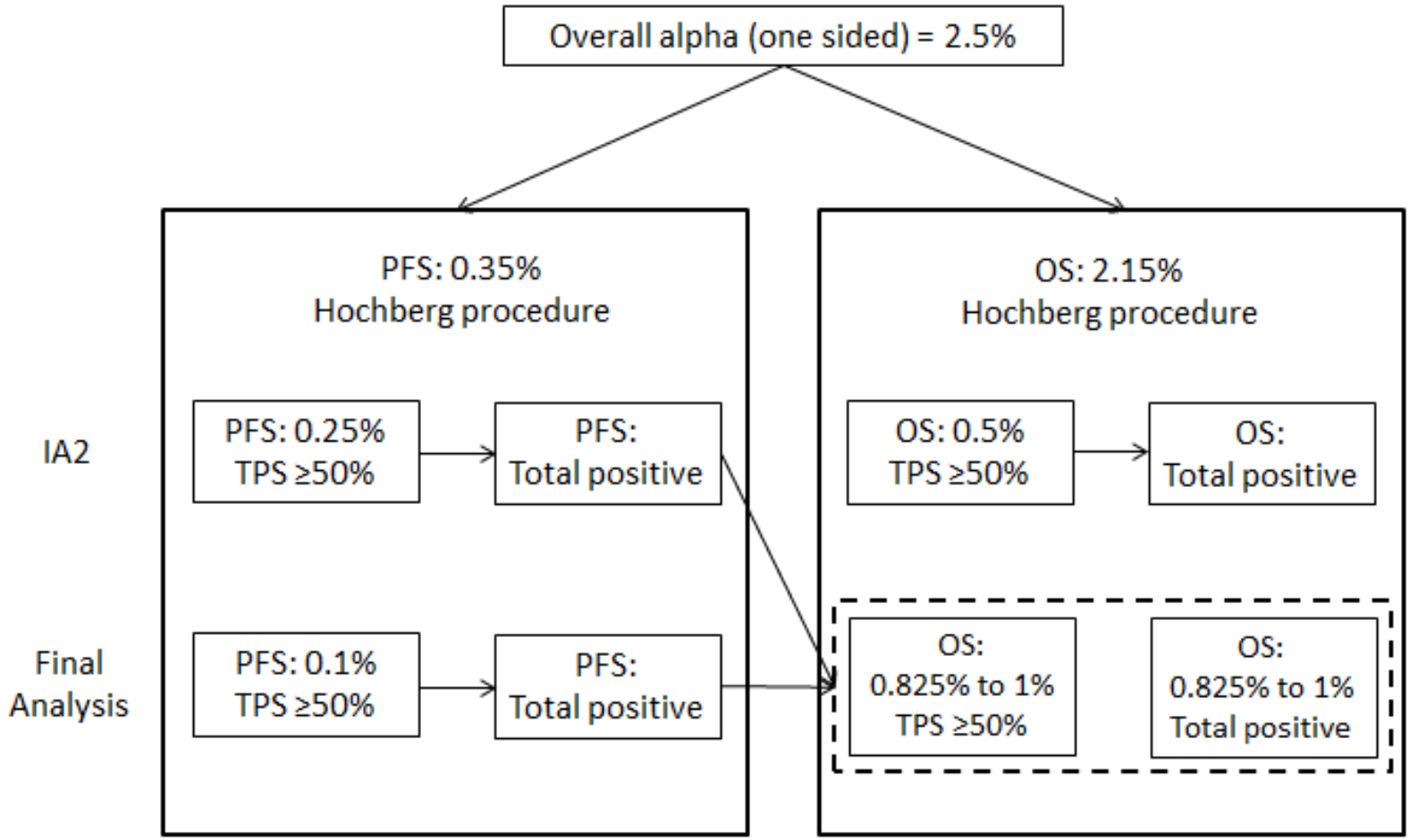


R = Randomization PD = Progressive Disease SFU = Survival Follow-up

Herbst et al. [2016]

Keynote 010 Statistical Methods: Multiplicity

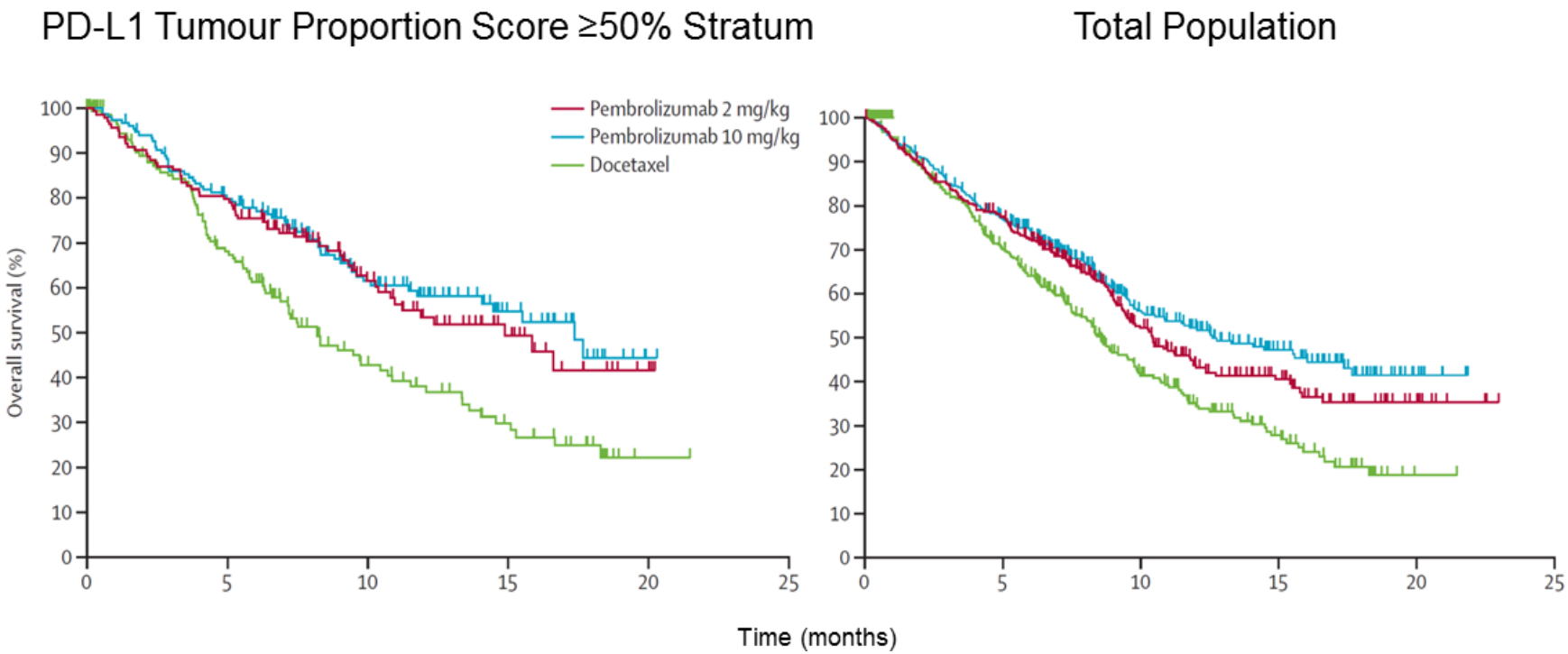
Type I error divided and reallocated between statistical tests



MERCK

Herbst et al. [2016]

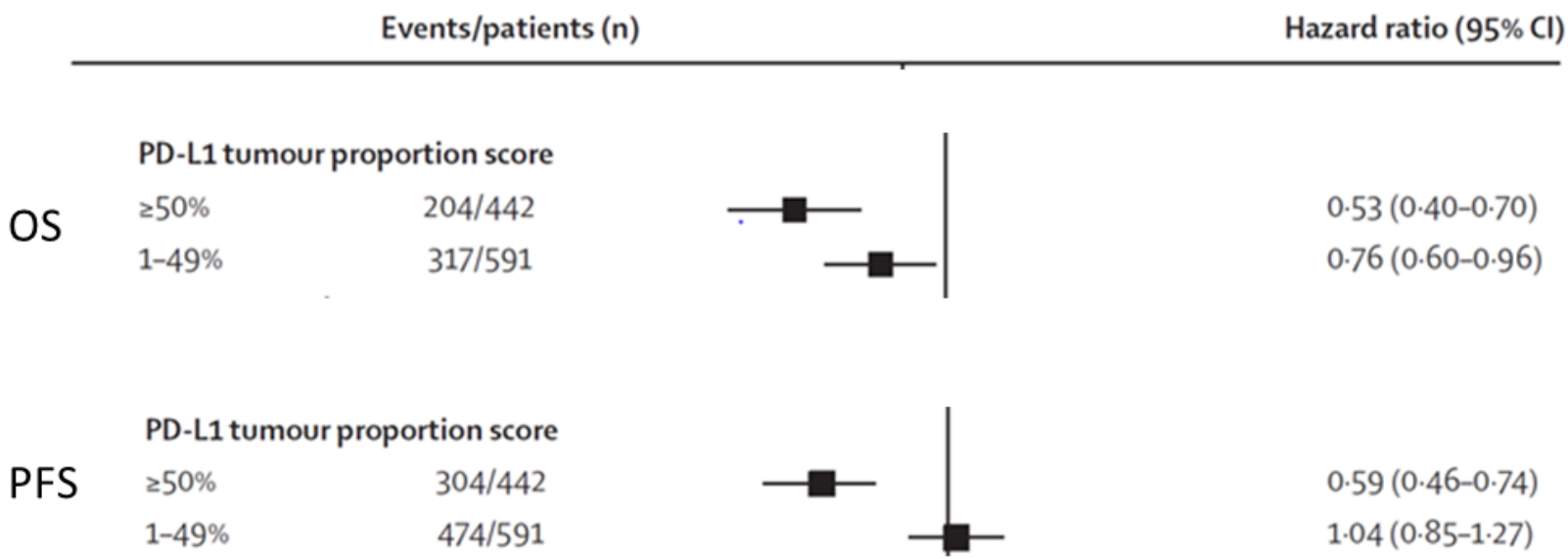
Keynote 010: Overall Survival (OS) by Biomarker at IA2



OS achieved statistical significance according to multiplicity plan in both treatment groups, total positive and TPS $\geq$ 50% populations
Herbst et al. [2016]; PFS=progression free survival

MERCK

Keynote 010: OS and PFS Hazard Ratio by Biomarker



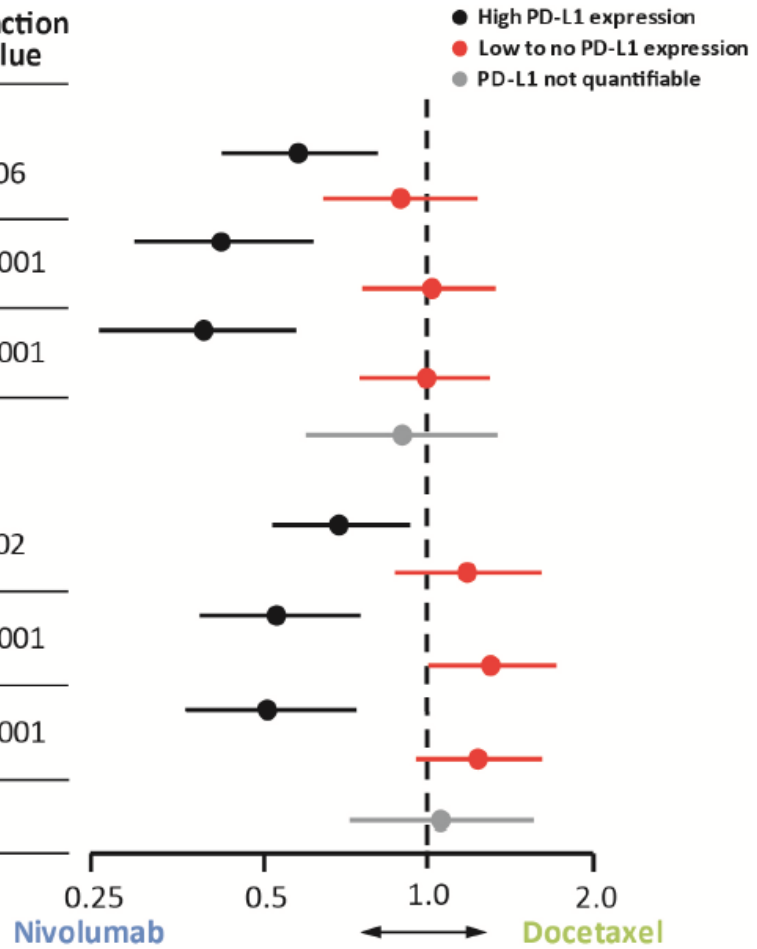
Herbst et al. [2016]

MERCK

CheckMate 57 Efficacy by Biomarker Status (Exploratory)

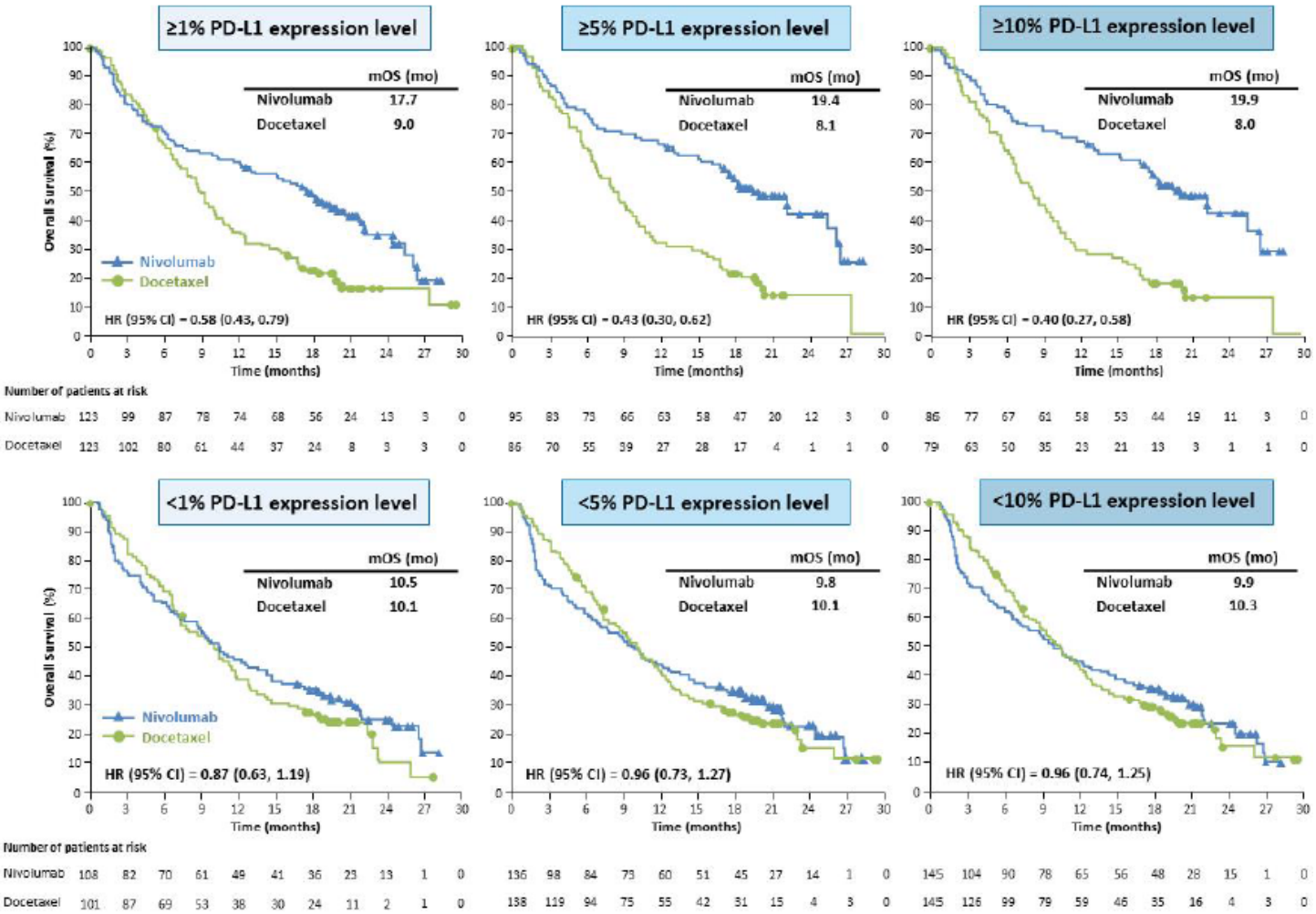
Primary evaluations and approval in broad population

PD-L1 expression level	Nivolumab n	Docetaxel n	Unstratified HR (95% CI)	Interaction *P*-value
OS				
≥1%	123	123	0.59 (0.43, 0.82)	0.06
<1%	108	101	0.90 (0.66, 1.24)	
≥5%	95	86	0.43 (0.30, 0.63)	<0.001
<5%	136	138	1.01 (0.77, 1.34)	
≥10%	86	79	0.40 (0.26, 0.59)	<0.001
<10%	145	145	1.00 (0.76, 1.31)	
Not quantifiable at baseline	61	66	0.91 (0.61, 1.35)	
PFS				
≥1%	123	123	0.70 (0.53, 0.94)	0.02
<1%	108	101	1.19 (0.88, 1.61)	
≥5%	95	86	0.54 (0.39, 0.76)	<0.001
<5%	136	138	1.31 (1.01, 1.71)	
≥10%	86	79	0.52 (0.37, 0.75)	<0.001
<10%	145	145	1.24 (0.96, 1.61)	
Not quantifiable at baseline	61	66	1.06 (0.73, 1.56)	



Borghaei et al. [2015], supplementary materials

CheckMate 57 OS by Biomarker Status (Exploratory)



Borghaei et al. [2015], supplementary materials

MERCK

PD-L1 Biomarker is Predictive

- Previously treated (non-squamous) NSCLC
- For high PD-L1 measures, PD-1 antibodies (nivolumab, pembrolizumab) are effective
- 'Exact' cutoffs for effectiveness unknown
- Companion diagnostic approach (pembrolizumab)
  - Trial and approval in targeted-population only
  - Possibility of patient benefit in broad population?
  - Likely higher probability of success from start

MERCK

PD-L1 Biomarker is Predictive

- Previously treated (non-squamous) NSCLC
- For high PD-L1 measures, PD-1 antibodies (nivolumab, pembrolizumab) are effective
- 'Exact' cutoffs for effectiveness unknown
- Companion diagnostic approach (pembrolizumab)
  - Trial and approval in targeted-population only
  - Possibility of patient benefit in broad population?
  - Likely higher probability of success from start
- Complementary diagnostic approach (nivolumab)
  - No α-controlled subgroup testing
  - Approval in broad population
  - Diagnostic approved (first 'complementary' diagnostic)
  - Risk that overall population could have not demonstrated statistical significance

MERCK

PD-L1 Biomarker is Predictive

- Previously treated (non-squamous) NSCLC
- For high PD-L1 measures, PD-1 antibodies (nivolumab, pembrolizumab) are effective
- 'Exact' cutoffs for effectiveness unknown

- Complementary diagnostic approach (nivolumab)
  - No α-controlled subgroup testing
  - Approval in broad population
  - Diagnostic approved (first 'complementary' diagnostic)
  - Risk that overall population could have not demonstrated statistical significance

MERCK

Multiple histology, biomarker focused, single arm trials

- Key endpoints: response rate and duration of response
- Phase IB studies
  - Keynote 12: A Phase Ib Multi-Cohort Study of MK-3475 in Subjects With Advanced Solid Tumors
    - Cancer types: breast (triple negative), head and neck cancer, urothelial tract, gastric
  - Keynote 28: Phase IB Study of Pembrolizumab (MK-3475) in Subjects With Select Advanced Solid Tumors

MERCK

Multiple histology, biomarker focused, single arm trials

- Phase IB studies
  - Keynote 12: A Phase Ib Multi-Cohort Study of MK-3475 in Subjects With Advanced Solid Tumors
    - Cancer types: breast (triple negative), head and neck cancer, urothelial tract, gastric
  - Keynote 28: Phase IB Study of Pembrolizumab (MK-3475) in Subjects With Select Advanced Solid Tumors
- Keynote 158: A Clinical Trial of Pembrolizumab (MK-3475) Evaluating Predictive Biomarkers in Subjects With Advanced Solid Tumors
  - Cohorts A-J: 10 solid tumor types (overlap with Keynote 12, 28)
  - Cohort K: MSI-high (biomarker) selected population, any solid tumor type
  - N=1100

MERCK

Multiple histology, biomarker focused, single arm trials

- Key endpoints: response rate and duration of response

- Keynote 158: A Clinical Trial of Pembrolizumab (MK-3475) Evaluating Predictive Biomarkers in Subjects With Advanced Solid Tumors
  - Cohorts A-J: 10 solid tumor types (overlap with Keynote 12, 28)
  - Cohort K: MSI-high (biomarker) selected population, any solid tumor type
  - N=1100

MERCK

Strong Type I error control for overall population and biomarker subgroups in randomized studies

MERCK

Hypotheses, analyses and test statistics

- $h > 1$ hypotheses
- $k > 1$ analyses
- $T_1 < T_2 \cdots < T_k$ calendar times of analyses
- For hypothesis $i = 1, \ldots, h$
  - May not test at all times for each hypothesis
  - $1 \leq k(i) \leq k$ analysis times $T_{i,1} < T_{i,2} \cdots < T_{i,k(i)}$
    - Some or all of $T_1 < T_2 \cdots < T_k$
  - $d_{i,1} < d_{i,2} \cdots < d_{i,k(i)}$ events for each analysis
  - $\mathcal{I}_{i,1} < \mathcal{I}_{i,2} \cdots < \mathcal{I}_{i,k(i)}$ statistical information for each analysis
    - For equal randomization with a time-to-event endpoint, this is approximated by $d_{i,j}/4$ [Schoenfeld, 1981]
  - $Z_{i,1}, Z_{i,2}, \ldots, Z_{i,k(i)}$ group sequential, normal test statistics with variance 1 to test hypothesis

MERCK

Example 1: Calendar-based

- $h = 2$ hypotheses (2 endpoints)
  - PFS: progression free survival; time until progression or death
  - OS: overall survival; time until death
- $k = 3$ analyses at given calendar times
  - PFS analyzed at $T_1 = 18$ and $T_2 = 24$ months
  - OS analyzed at $T_1 = 18$, $T_2 = 24$ and $T_3 = 36$ months
  - Number of events (d_{ij}) is random

MERCK

Example 2: Event-based

- $h = 2$, same hypotheses
  - PFS
  - OS
- $k = 3$ analyses at 3 times
  - PFS analyzed twice after $d_{1,1} < d_{1,2}$ endpoints
  - OS analyzed at same times plus final analysis with $d_{2,3}$ endpoints
  - Number of OS events at interims $(d_{2,1}, d_{2,2})$ are random
  - Analysis times $T_1,\ T_2,\ T_3$ are random

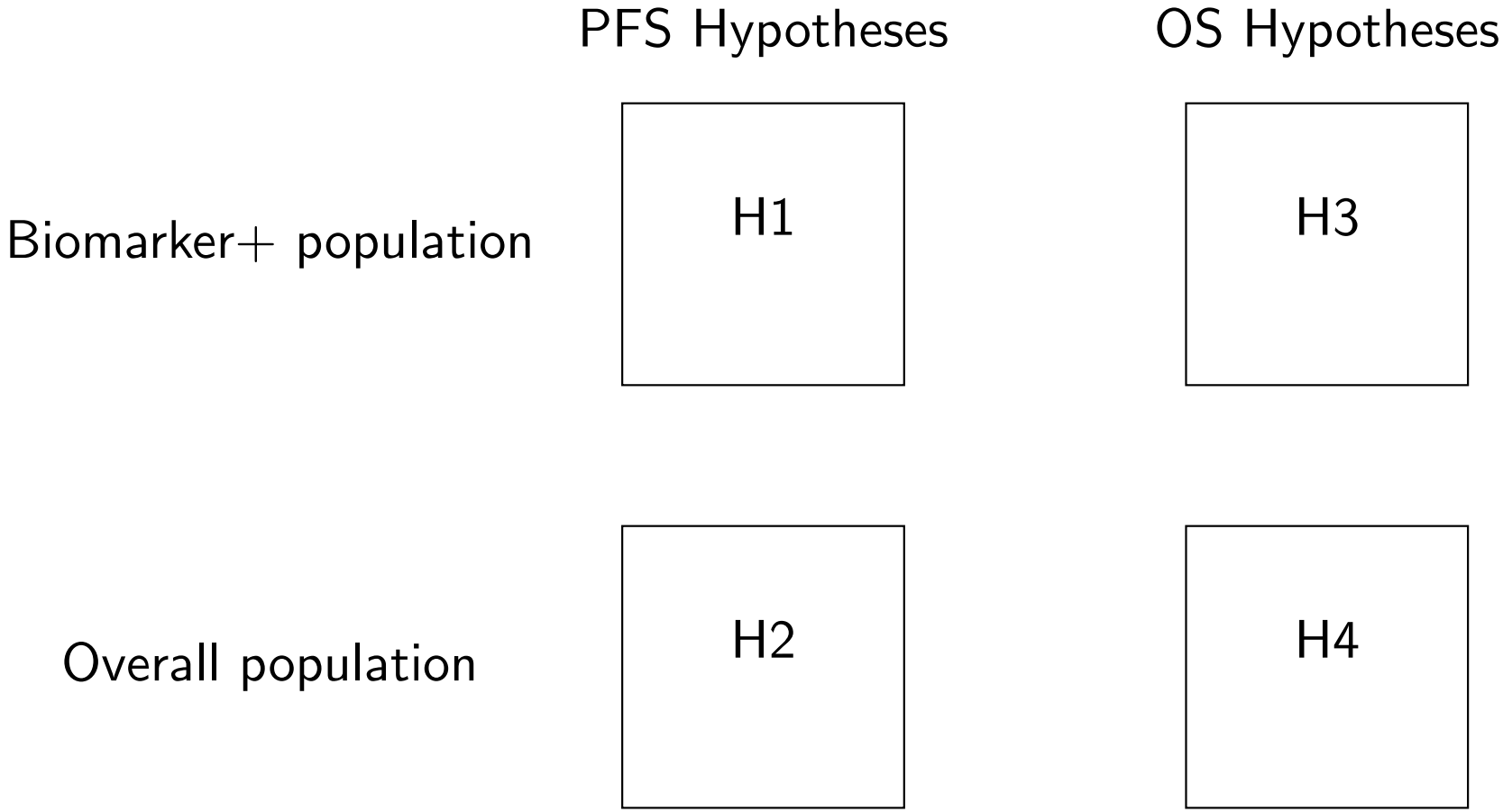
MERCK

Example 3: Add biomarker hypothesis

- $h = 4$ hypotheses (2 endpoints $\times$ 2 populations)
  - PFS
    - H_1: BM+ population (biomarker positive subgroup)
    - H_2: Overall population
  - OS
    - H_3: BM+ population
    - H_4: Overall population
- $k = 3$ analyses
  - PFS analyzed twice after $d_{1,1} < d_{1,2}$ endpoints
  - PFS endpoints for overall population $(d_{2,1}, d_{2,2})$ are random
  - OS analyzed at same times plus final analysis with $d_{3,3}$ endpoints
  - Number of OS events are random for BM+ at interims $(d_{3,1}, d_{3,2})$
  - Number of OS events are random for total population at all analyses $(d_{4,1}, d_{4,2}, d_{4,3})$
  - Analysis times T_1, T_2, T_3 are random

MERCK

Example 3 graph



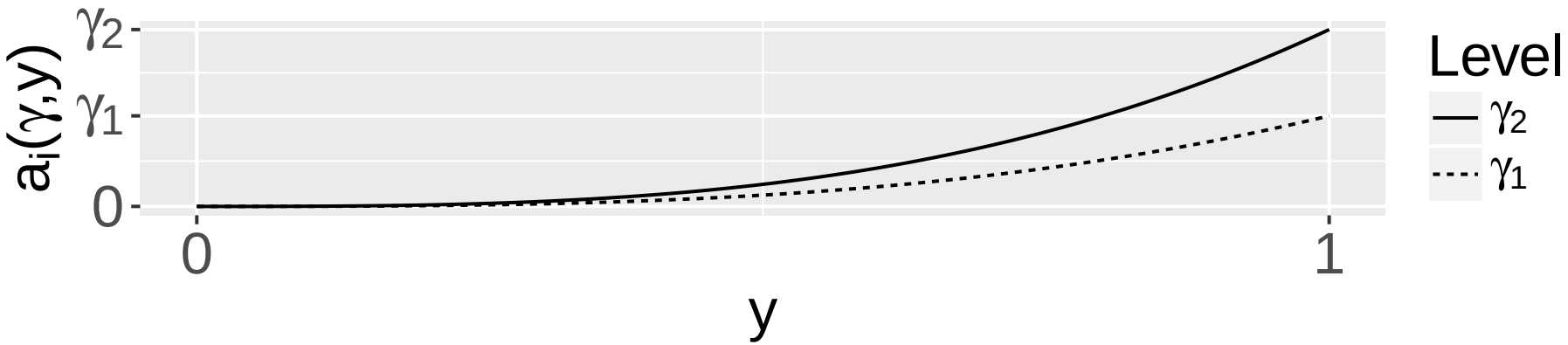
Methods summary

- Spending time concept importance when multiple hypotheses tested with varying rates of information accumulation
- Some key concepts for spending functions with group sequential combined with graphical multiplicity [Maurer and Bretz, 2013]
- Brief note on technical algorithm for testing

MERCK

Spending functions

- Slightly modified notation of Maurer and Bretz [2013]
- All hypotheses controlled at 1-sided level $0 < \alpha < 1$
- Spending function for hypothesis i
  - $a_i(\gamma, y)$
  - Non-decreasing for y on $[0, 1]$, γ on $[0, \alpha]$
  - $a_i(\gamma, 0) = 0$
  - $a_i(\gamma, y) = \gamma$ for $y = 1$



Spending times

For hypothesis $i = 1, \ldots, k$

- Test hypothesis at level γ_i
- $k(i) \leq k$ analysis times
- Spending times

$$0 = t_{i,0} < t_{i,1} \leq \cdots \leq t_{i,k(i)} = 1$$

- Type I error allocated to analysis $j = 1, \ldots, k(i)$

$$a(\gamma_i, t_{i,j}) - a(\gamma_i, t_{i,j-1})$$

- Bounds for Z-statistics then a standard group sequential calculation [Slud and Wei, 1982] based on statistical information (endpoint count; Tsiatis [1982])
- Any futility bound ignored in calculations per Liu and Anderson [2008]

Setting spending times

- $d_{i,\max}$: maximum planned endpoints for hypothesis i
- Information time [Lan and DeMets, 1983]:

$$t_{i,j} = \min(1, \mathcal{I}_{i,j}/\mathcal{I}_{i,\max} = d_{i,j}/d_{i,\max})$$

  - Under-running variation is to set $t_{i,k(i)} = 1$ if planned information level not reached for hypothesis i ($\mathcal{I}_{i,k(i)} < \mathcal{I}_{i,\max}$)
- Calendar time [Lan and DeMets, 1989]:

$$t_{i,j} = T_{i,j}/T_{i,k(i)}$$

- Alternate information time:
  - *e.g.*, set all OS hypotheses based on information time for biomarker positive subgroup OS hypothesis
  - Suggested in Proschan et al. [2006] (Section 5.1.1) based on the same logic as calendar time

MERCK

Key condition for spending time

- In words: Cannot choose spending time or information time for an analysis based on value of current or former test statistic for any hypothesis.
- Using notation: $t_{i,j}$ and $\mathcal{I}_{i,j}$ are conditionally independent of $Z_{i',j'} - E\{Z_{i',j'}\}$ for $T_{i',j'} \leq T_{i,j}$, $i' = 1, 2, \ldots, k$.

MERCK

Testing algorithm

For a given analysis

1. Test each null hypothesis i to be tested at or before this analysis
   a) Analysis index j, denote Type I error allocated to hypothesis γ_i
   b) Compute spending time $t_{i,j}$
   c) Compute boundaries $b_{i,j'}$, $j' = 1, \ldots, j$ based on
      - γ_i
      - $t_{i,j'}$, $j' = 1, \ldots, j$
      - $\mathcal{I}_{i,j'}$, $j' = 1, \ldots, j$
      - This is a standard group sequential design calculation
      - For $j' < j$, $b_{i,j'}$ will not change unless γ_i has changed due to reallocation
2. If $Z_{i,j'} > b_{i,j'}$ for any $j' = 1, \ldots, j$, reject null hypothesis i
3. If any hypothesis was rejected, reallocate γ_i per multiplicity graph [Bretz et al., 2009] and return to step 1

MERCK

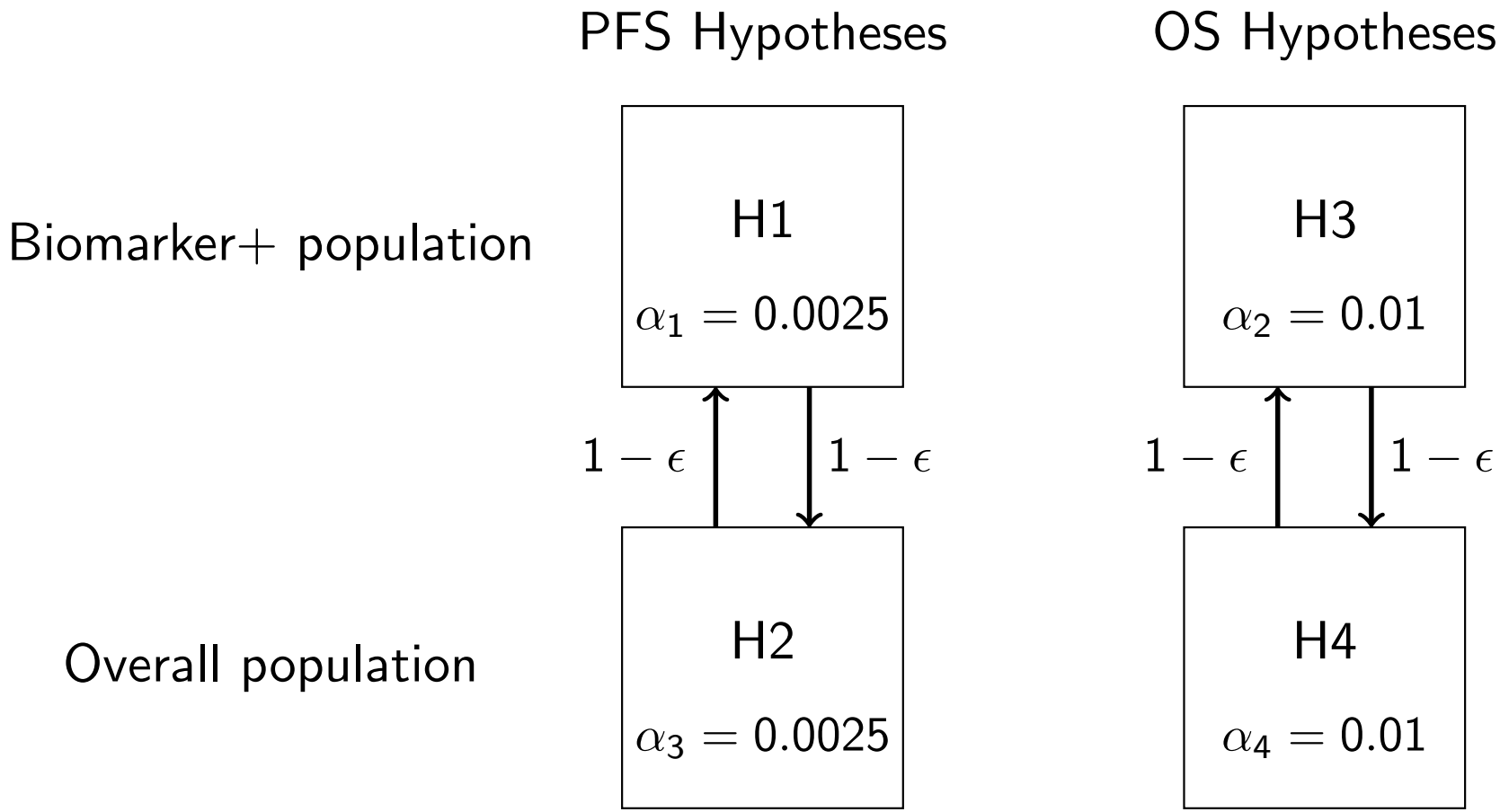
Return to historical example; Type I error allocation

	PFS Hypotheses	OS Hypotheses
Biomarker+ population	H1 $\alpha_1 = 0.0025$	H3 $\alpha_2 = 0.01$
Overall population	H2 $\alpha_3 = 0.0025$	H4 $\alpha_4 = 0.01$

- Allocate most α to OS
- Equal split between populations

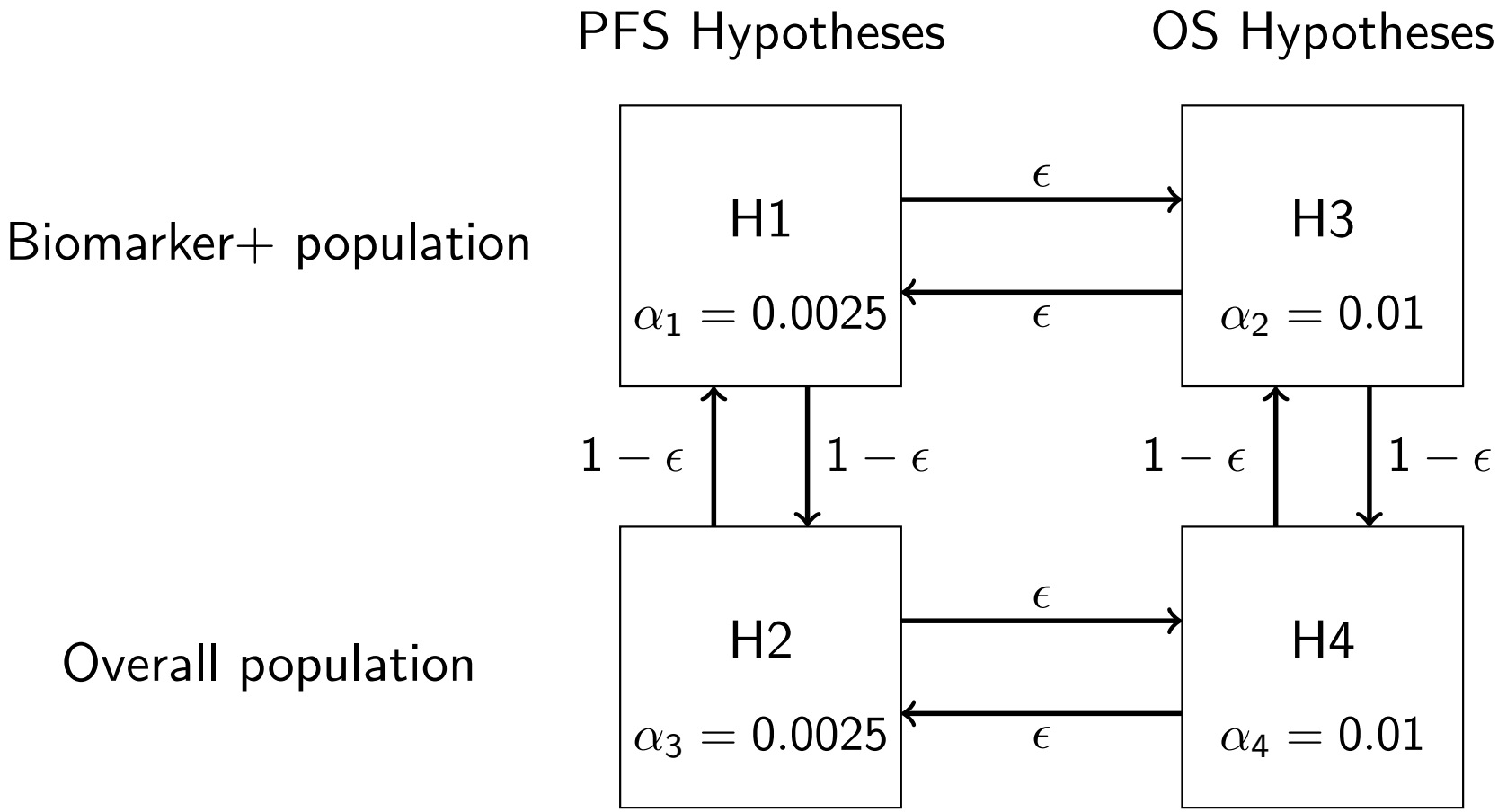
MERCK

Main reallocation within endpoints



- Bonferonni-Holm between populations
- Motivated by PFS testing ending before OS testing

Final reallocation between endpoints



- Bonferonni-Holm between populations
- If both populations reject, reallocate to other endpoint [Bretz et al., 2009]

Summary and conclusions

- Biomarker-based development programs continue to be of interest
- Different approaches have been taken
- Personalized medicine results and multiple endpoints of interest can create a substantial multiplicity problem for oncology development
- Maurer and Bretz [2013] creates a framework for group sequential trials with multiple hypotheses
- Method extended here to apply to trials with time-to-event endpoints with varying rates of information (endpoint) accumulation

MERCK

References

Hossein Borghaei, Luis Paz-Ares, Leora Horn, David R Spigel, Martin Steins, Neal E Ready, Laura Q Chow, Everett E Vokes, Enriqueta Felip, Esther Holgado, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *New England Journal of Medicine*, 373(17):1627–1639, 2015.

Frank Bretz, Willi Maurer, and Martin Posch. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*, 28:586–604, 2009. doi: 10.1002/sim.3495.

Edward B Garon, Naiyer A Rizvi, Rina Hui, Natasha Leighl, Ani S Balmanoukian, Joseph Paul Eder, Amita Patnaik, Charu Aggarwal, Matthew Gubens, Leora Horn, et al. Pembrolizumab for the treatment of non–small-cell lung cancer. *New England Journal of Medicine*, 372(21):2018–2028, 2015.

Roy S Herbst, Paul Baas, Dong-Wan Kim, Enriqueta Felip, José L Pérez-Gracia, Ji-Youn Han, Julian Molina, Joo-Hang Kim, Catherine Dubos Arvis, Myung-Ju Ahn, et al. Pembrolizumab versus docetaxel for previously treated, pd-l1-positive, advanced non-small-cell lung cancer (keynote-010): a randomised controlled trial. *The Lancet*, 387(10027):1540–1550, 2016.

K. K. G. Lan and David L. DeMets. Discrete sequential boundaries for clinical trials. *Biometrika*, 70:659–663, 1983.

K. K. G. Lan and David L. DeMets. Group sequential procedures: Calendar versus information time. *Statistics in Medicine*, 8: 1191–1198, 1989. doi: 10.1002/sim.4780081003.

Qing Liu and Keaven M. Anderson. On adaptive extensions of group sequential trials for clinical investigations. *Journal of the American Statistical Association*, 103:1621–1630, 2008. doi: 10.1198/016214508000000986.

Willi Maurer and Frank Bretz. Multiple testing in group sequential trials using graphical approaches. *Statistics in Biopharmaceutical Research*, 5:311–320, 2013. doi: 10.1080/19466315.2013.807748.

Michael A. Proschan, K. K. Gordon Lan, and Janet Turk Wittes. *Statistical Monitoring of Clinical Trials. A Unified Approach.* Springer, New York, NY, 2006.

David Schoenfeld. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68: 316–319, 1981.

Eric V. Slud and L. J. Wei. Two-sample repeated significance tests based on the modified wilcoxon statistic. *Journal of the American Statistical Association*, 77:862–868, 1982.

Anastasios A. Tsiatis. Repeated significance testing for a general class of statistics use in censored survival analysis. *Journal of the American Statistical Association*, 77:855–861, 1982.