Interim Futility Monitoring When Assessing Immune Therapies With A Potentially Delayed Treatment Effect

Boris Freidlin
Edward Korn
National Cancer Institute
Bethesda, MD

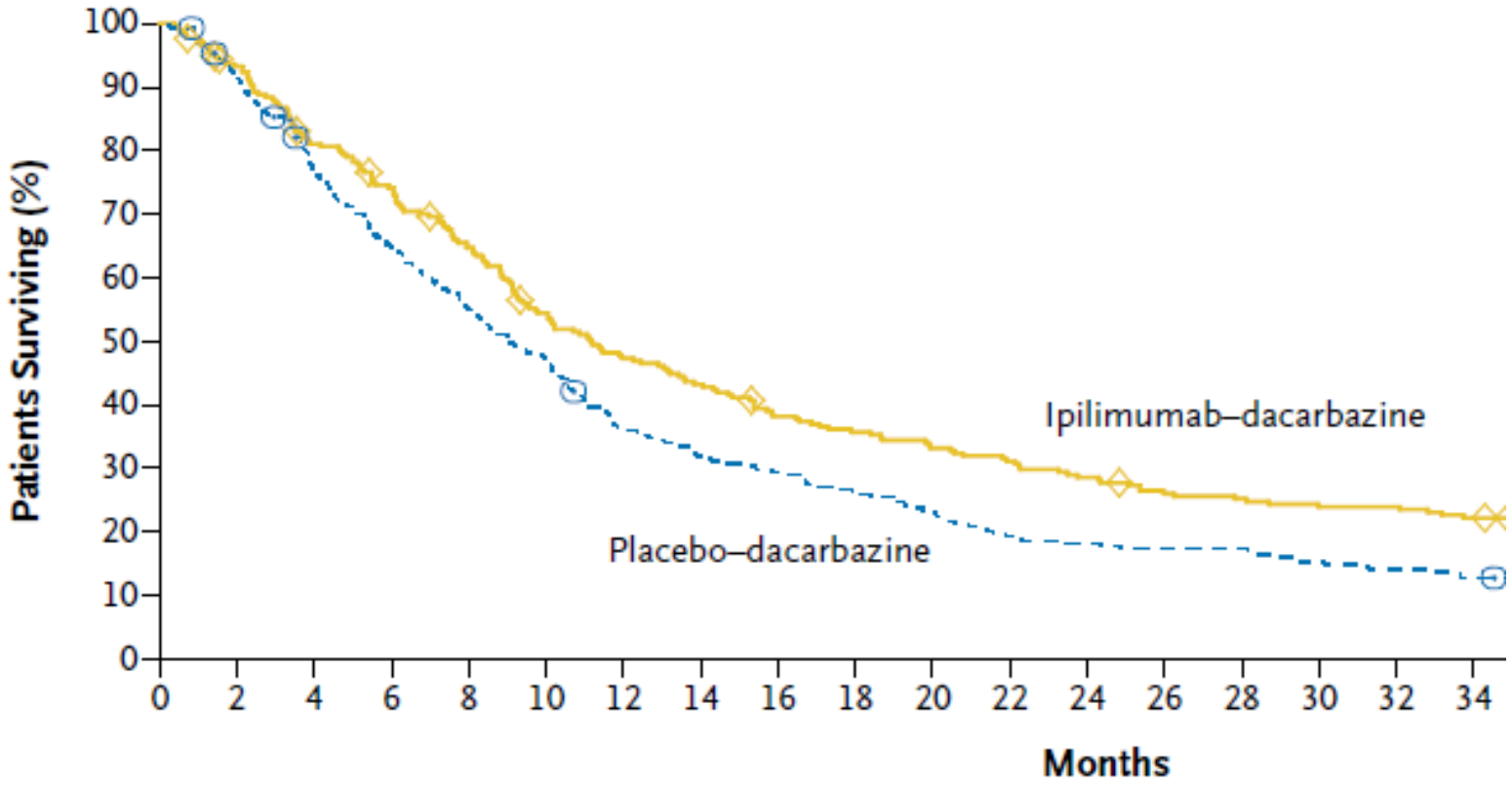
Motivation

- Introduction of new immune therapies that may have a delayed treatment effect necessitates re-evaluation of traditional clinical trial designs in oncology.

- A key feature of RCT is interim futility monitoring which protects patients and resources if the experimental treatment is detrimental or unlikely to be shown superior to the standard treatment.

- The appropriateness of futility monitoring is frequently questioned when the effect of the experimental treatment may be delayed, e.g., in trials of immune agents.

Delayed treatment effect

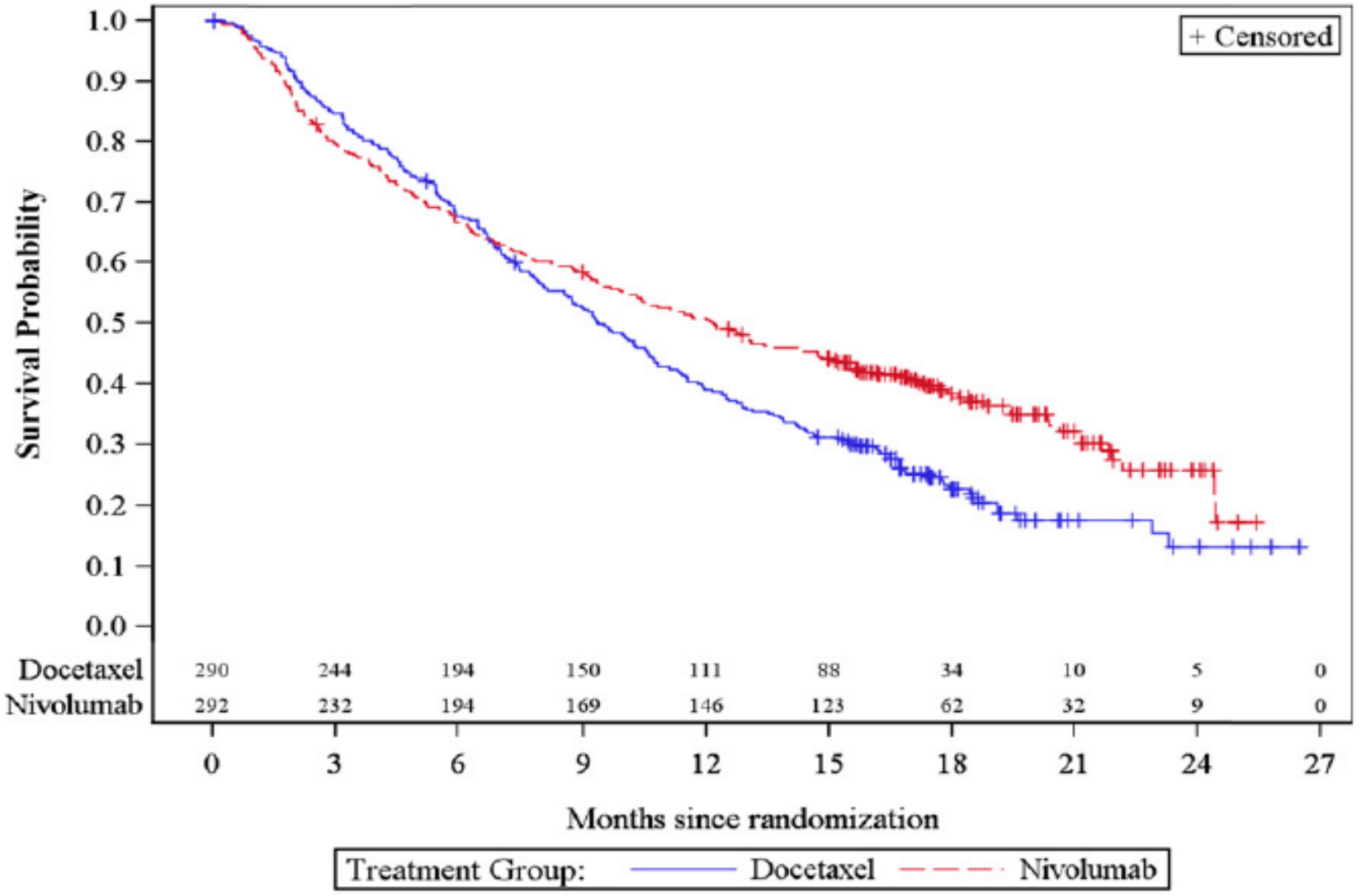


No. at Risk

Ipilimumab–dacarbazine	250	230	199	181	157	131	114	104	91	85	79	74	68	61	59	56	56	52
Placebo–dacarbazine	252	229	190	160	136	116	89	78	72	64	56	47	44	42	42	37	34	31

Robert et al NEJM 2011: **Melanoma**

Crossing hazards, crossing survival curves



CheckMate 057 NEJM 2015: **Non-squamous NSCLC**

RCT designs:

Possible

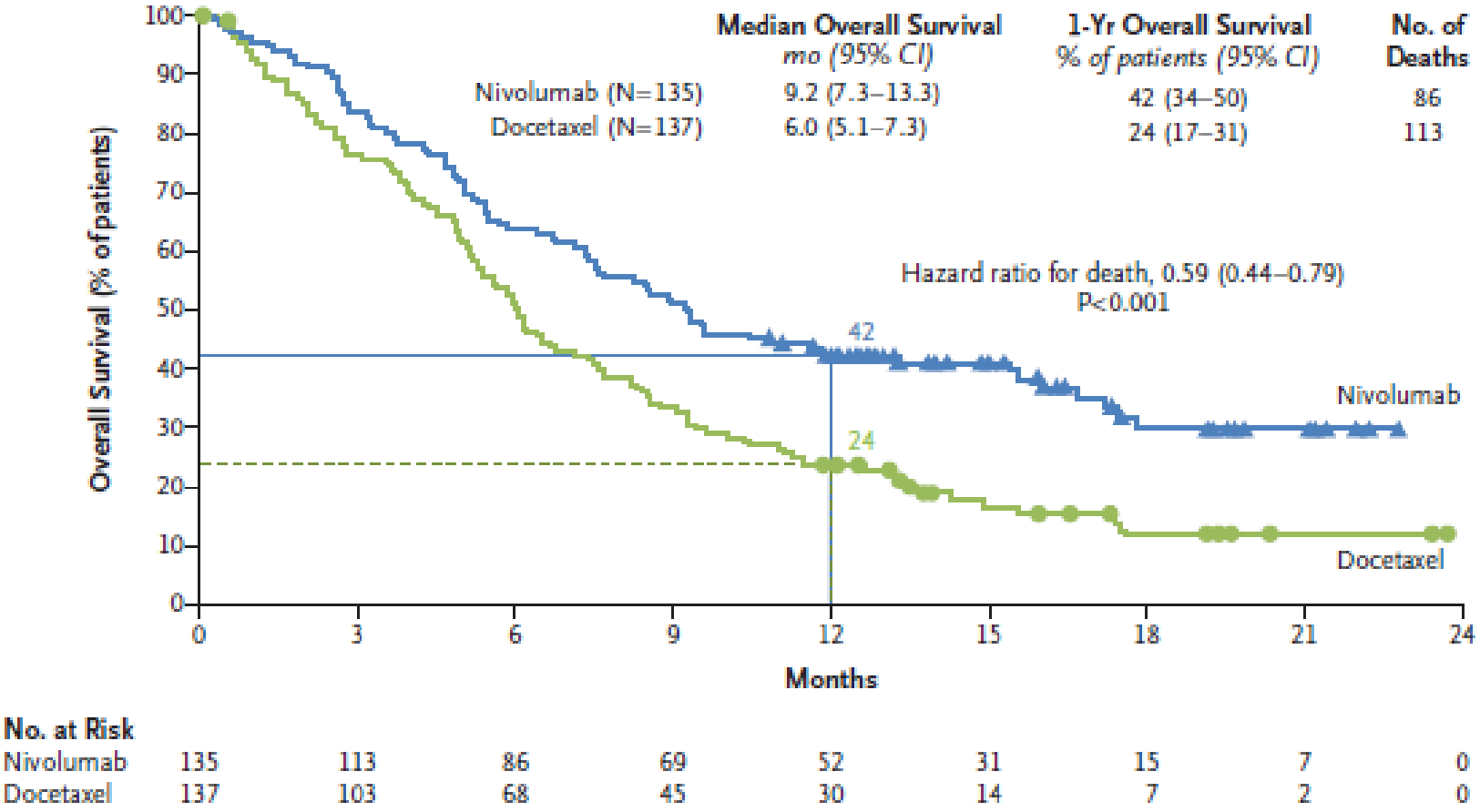
Experimental vs Active control

Unlikely

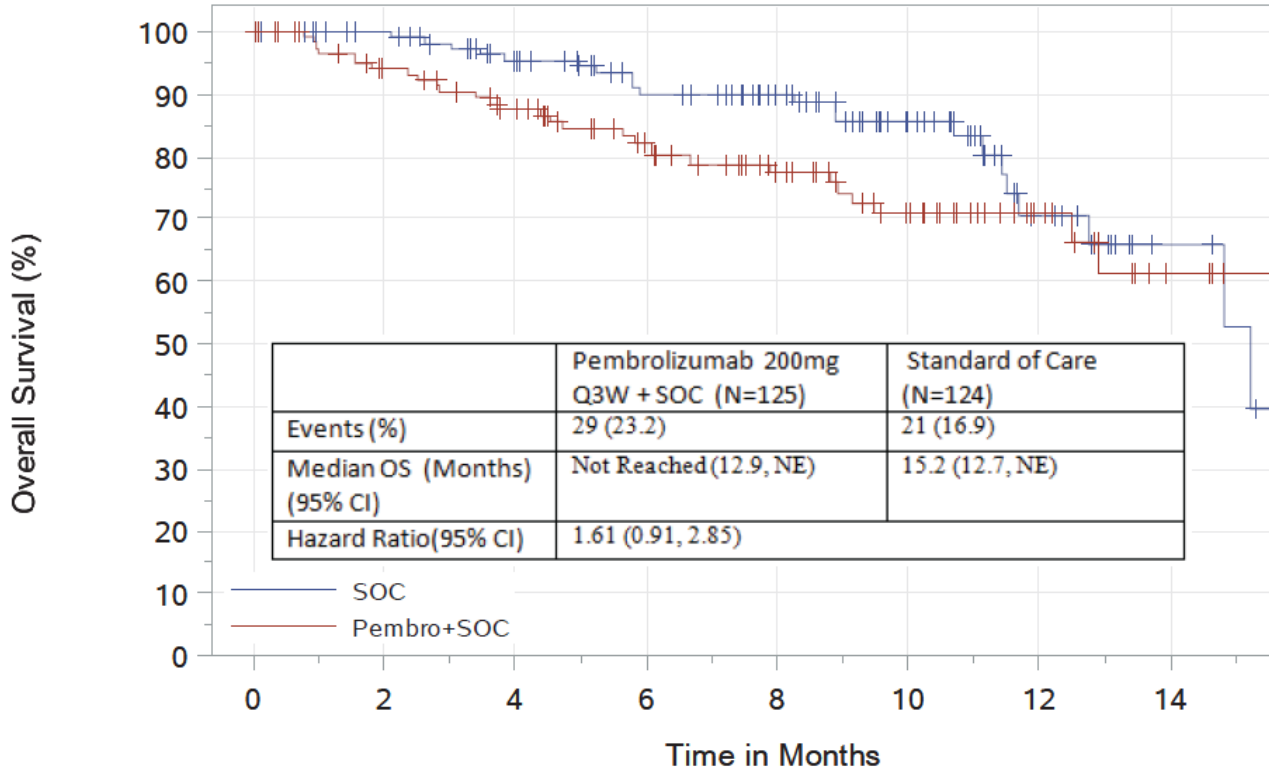
Experimental + Active vs Active

Or

Experimental vs Placebo



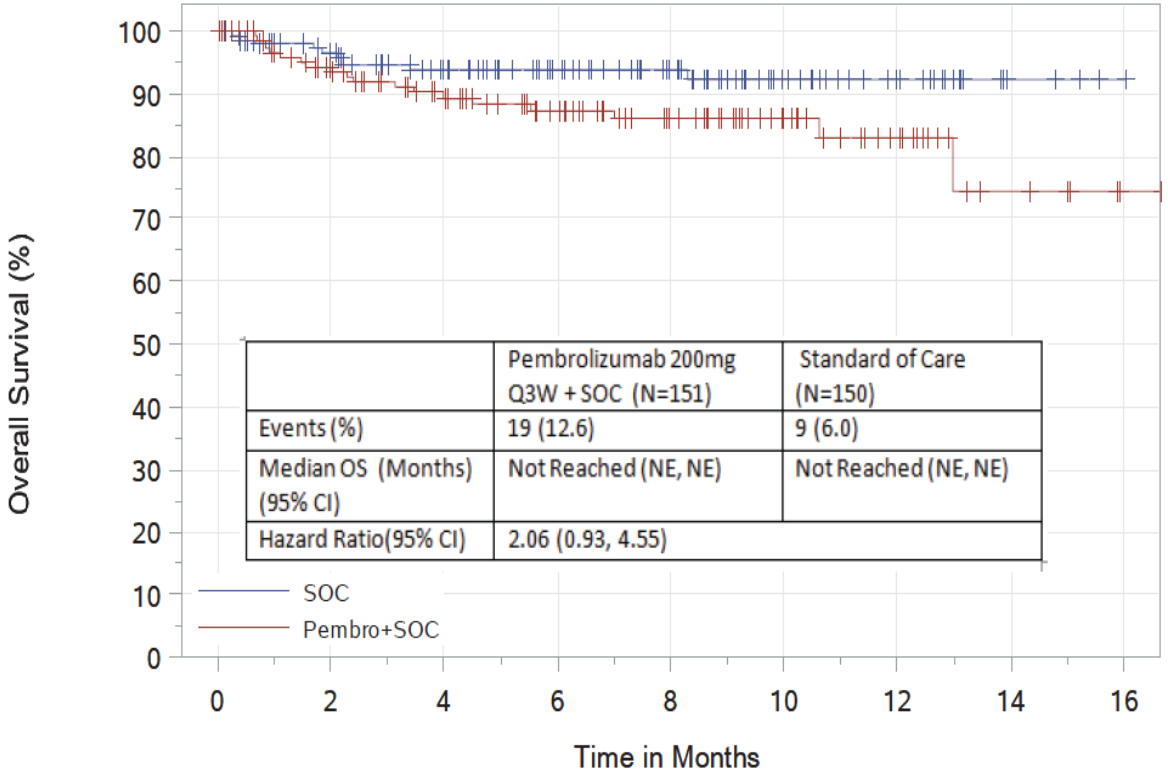
Brahmer et al NEJM 2015: **Squamous NSCLC**



	Pembrolizumab 200mg Q3W + SOC (N=125)	Standard of Care (N=124)
Events (%)	29 (23.2)	21 (16.9)
Median OS (Months) (95% CI)	Not Reached (12.9, NE)	15.2 (12.7, NE)
Hazard Ratio(95% CI)	1.61 (0.91, 2.85)	

Number of Subjects at Risk

SOC	124	115	99	83	67	42	18	6
Pembro+SOC	125	105	91	73	53	37	18	7



	Pembrolizumab 200mg Q3W + SOC (N=151)	Standard of Care (N=150)
Events (%)	19 (12.6)	9 (6.0)
Median OS (Months) (95% CI)	Not Reached (NE, NE)	Not Reached (NE, NE)
Hazard Ratio(95% CI)	2.06 (0.93, 4.55)	

Number of Subjects at Risk

SOC	150	124	102	82	56	31	19	5	1
Pembro+SOC	151	122	100	79	58	32	20	7	2

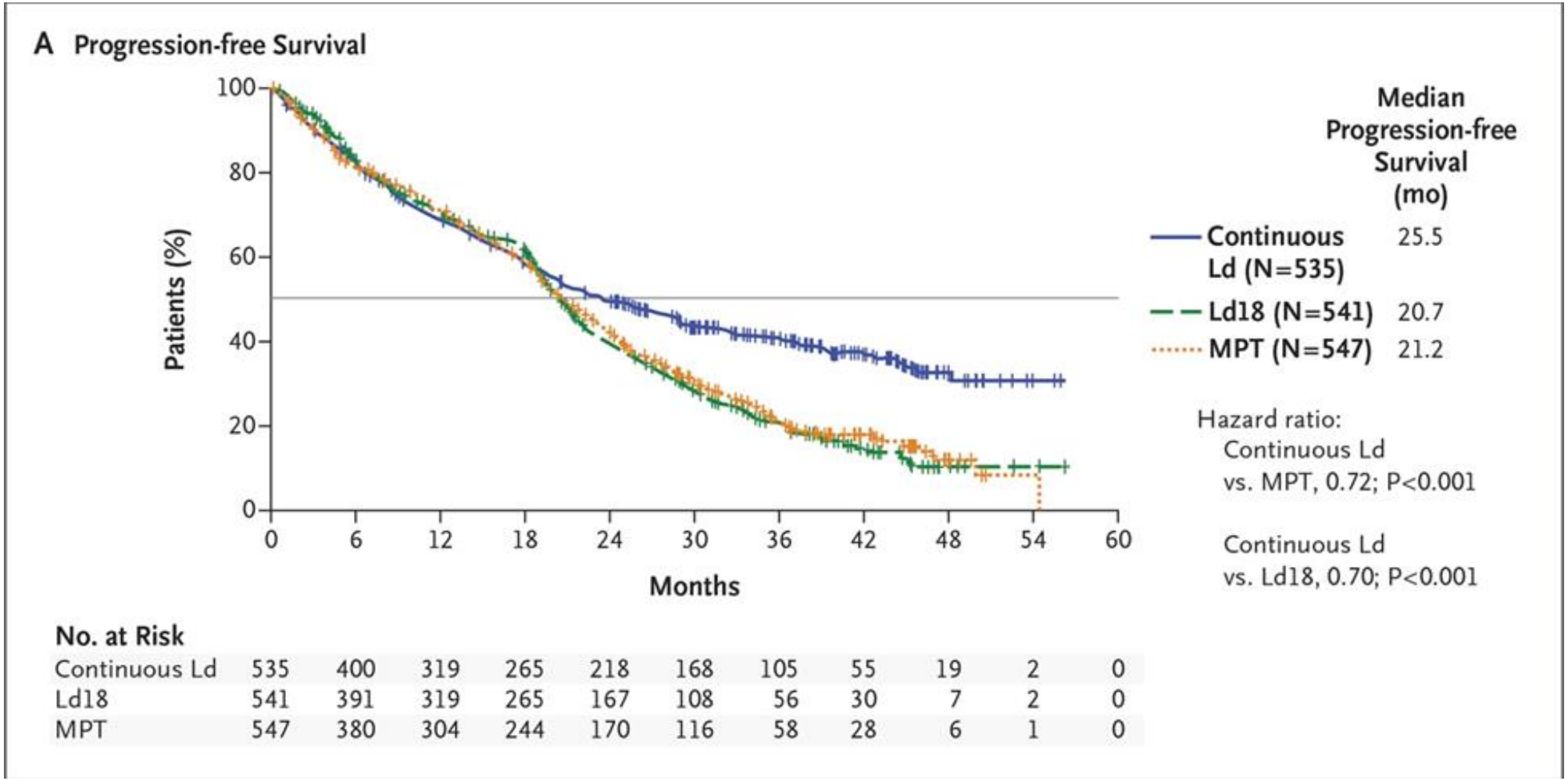
FDA Alerts: https://www.fda.gov/Drugs/DrugSafety/ucm574305.htm: **Myeloma**

Selecting a futility rule requires balancing patient safety and public health considerations:

- If the experimental therapy is ineffective, one would want to minimize the number of patients exposed to the therapy.

- If the new therapy is beneficial one would want to maximize the probability of detecting the benefit.

Aside: Delayed treatment effect built into trial design



FIRST trial Benboubker NEJM 2014: **Myeloma**

Goals

- Examine benefits and risks of using common futility monitoring approaches when there is a delayed treatment effect.

- Develop a new futility monitoring rule for use when there is a potential delayed treatment effect in an immunotherapy trial

Common futility approaches considered:

1) Wieand rule (Wieand Stat Med 1994):

First interim analysis (50% of total expected events): stop if observed HR >1

Second interim analysis (75% of total expected events): stop if observed HR >1

2) O'Brien-Fleming β-spending function (power family, Wang-Tsiatis, Pampallona-Tsiatis):

First interim analysis (33% of total expected events): stop if $Z<0.011$

Second interim analysis (66% of total expected events): stop if $Z<0.864$

(Tremelimumab in melanoma, Ribas et al NEJM 2013)

The proposed approach

A modification of the Wieand approach stop if observed HR>1: (expected treatment effect delay is **3-6 months**)*

First interim analysis: when at least 50% of the expected events have occurred **AND** at least two-thirds of the observed events have occurred later than **3 months** from randomization.

Second interim analysis: when at least 75% of the expected events have occurred **AND** at least two-thirds of the observed events have occurred later than **3 months** from randomization

* The rule is easily adjusted for longer expected delay periods

Reference: Korn and Freidlin, *Journal of Clinical Oncology*, 2018 p 2444-9

Simulated settings, 90% power without futility monitoring

- 680 patients randomized over **34 months**; the final analysis at 512 events.

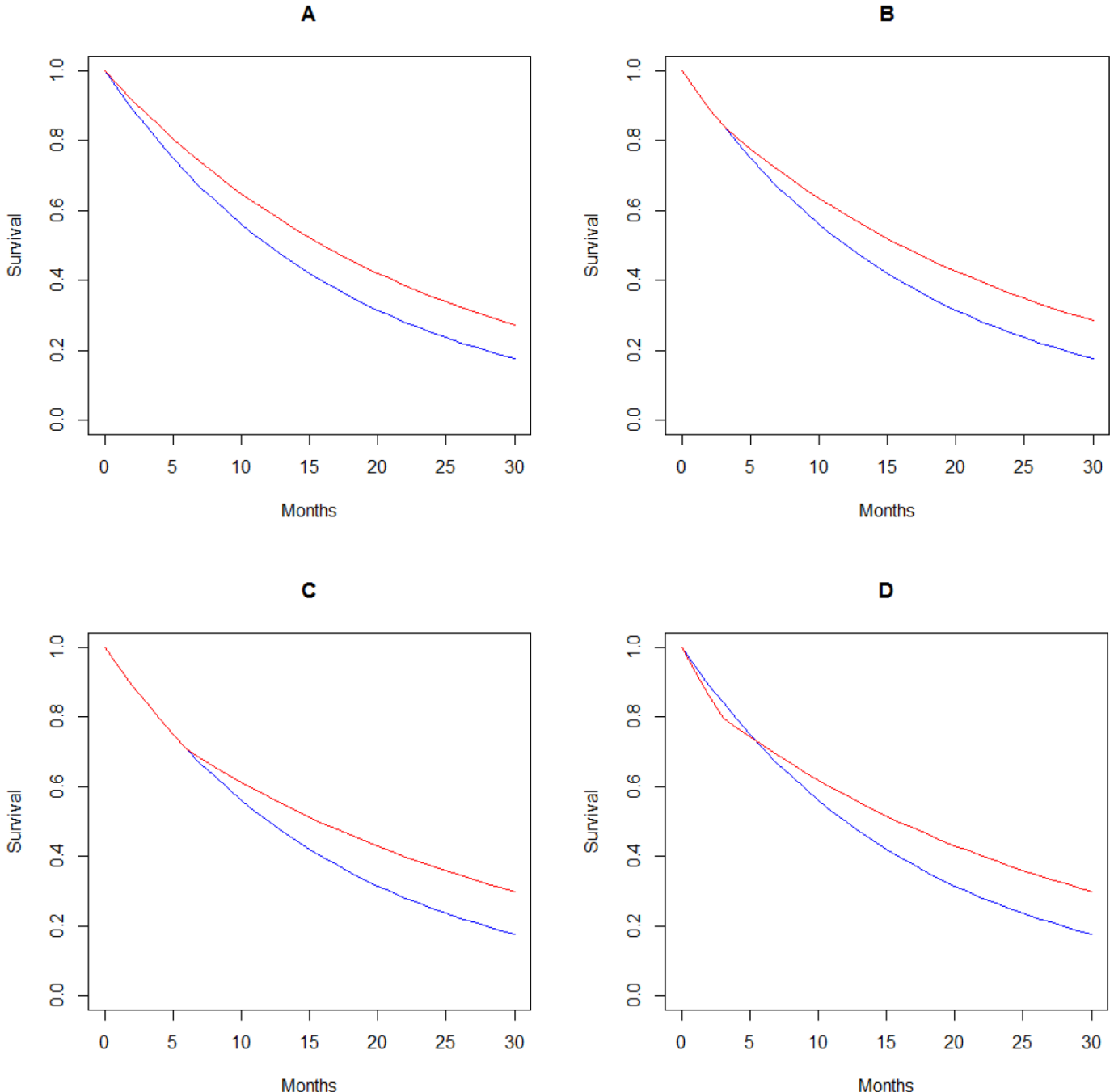
- 680 patients randomized over **12 months**; the final analysis at 512 events.

<u>Cure model (20% cure rate on the control arm)</u>

- 800 patients randomized over **34 months**; the final analysis at 512 events.

- 800 patients randomized over **12 months**; the final analysis at 512 events.

Scenarios considered



Korn and Freidlin, *Journal of Clinical Oncology*, 2018

680 patients accrued over **34 months**

Setting	No interim analyses		Standard Wieand		Standard O'Brien-Fleming		Proposed approach	
	Power	Duration SS	Power	Duration SS	Power	Duration SS	Power	Duration SS
HR=0.75 (no delay)	.900	47.4 680	.898	47.1 678.6	.879	46.2 671.7	.898	47.1 678.6
HR=1.00 (no delay)	.0247	43.9 680	.0245	34.1 602.6	.0226	28.2 529.1	.0245	34.3 605.3
HR=1.30 (no delay)	0	41.3 680	0	25.2 504.4	0	19.9 399.3	0	26.1 526.1
3 m delay	.901	47.8 680	.894	47.3 676.7	.846	45.4 661.2	.895	47.4 677.4
6 m delay	.903	48.3 680	.881	47.2 673.1	.786	43.9 644.3	.885	47.4 674.4
Crossing hazards	.899	48.3 680	.872	47.0 671.7	.762	43.2 638.0	.891	47.8 677.8

680 patients accrued over **12 months**

Setting	No interim analyses		Standard Wieand		Standard O'Brien-Fleming		Proposed approach	
	Power	Duration SS	Power	Duration SS	Power	Duration SS	Power	Duration SS
HR=0.75 (no delay)	.899	34.3 680	.897	34.1 680	.879	33.3 679.7	.898	34.2 680
HR=1.00 (no delay)	.0246	30.5 680	.0243	21.5 680	.0223	16.9 660.5	.0245	23.3 680
HR=1.30 (no delay)	0	27.7 680	0	13.7 680	0	10.9 607.5	0	19.7 680
3 m delay	.901	34.8 680	.882	33.9 680	.794	31.1 677.3	.898	34.5 680
6 m delay	.903	35.4 680	.804	32.1 680	.626	27.1 670.4	.879	34.3 680
Crossing hazards	.899	35.4 680	.781	31.6 680	.537	24.7 665.1	.895	35.1 680

20% cure rate: 800 patients accrued over **34 months**

Setting	No interim analyses		Standard Wieand		Standard O'Brien-Fleming		Proposed approach	
	Power	Duration SS	Power	Duration SS	Power	Duration SS	Power	Duration SS
HR=0.75 (no delay)	.900	41.3 800	.898	41.1 797.5	.879	40.3 787.0	.898	41.1 797.7
HR=1.00 (no delay)	.0246	38.1 800	.0243	30.2 673.2	.023	25.2 576.8	.0244	30.7 682.3
HR=1.30 (no delay)	0	35.9 800	0	22.8 535.7	0	18.1 426.0	0	25.3 592.0
3 m delay	.902	41.7 800	.895	41.3 795.0	.850	39.8 773.4	.899	41.4 796.9
6 m delay	.902	42.2 800	.883	41.4 790.1	.798	38.7 751.2	.891	41.6 793.7
Crossing hazards	.901	42.2 800	.878	41.3 788.3	.778	38.3 743.4	.898	42.1 798.8

20% cure rate: 800 patients accrued over **12 months**

Setting	No interim analyses		Standard Wieand		Standard O'Brien-Fleming		Proposed approach	
	Power	Duration SS	Power	Duration SS	Power	Duration SS	Power	Duration SS
HR=0.75 (no delay)	.900	26.7 800	.898	26.6 800	.879	26.0 797.9	.900	26.6 800
HR=1.00 (no delay)	.0249	24.0 800	.0244	17.9 800	.0225	14.5 744.8	.0249	21.5 800
HR=1.30 (no delay)	0	21.9 800	0	12.3 796.1	0	10.0 655.9	0	21.3 800
3 m delay	.901	27.0 800	.875	26.2 800	.786	24.3 787.2	.901	27.0 800
6 m delay	.901	27.4 800	.778	24.7 800	.587	21.0 765.7	.894	27.1 800
Crossing hazards	.899	27.4 800	.731	24.0 800	.506	19.4 751.5	.899	27.4 800

Conclusions

- Commonly used futility rules are optimized for settings with no delay in the treatment effect.

- If the treatment effect is delayed, the application of many commonly used futility rules may result in loss of power because interim results are dominated by the early events.

- The proposed futility monitoring rule results in a very small loss of power regardless of whether the treatment effect is delayed (even with rapid accrual), but offers considerable savings in time and patients treated when the experimental treatment is no better than, or worse than, the standard treatment.

Boris Freidlin
NCI
freidlinb@ctep.nci.nih.gov
240-2766024