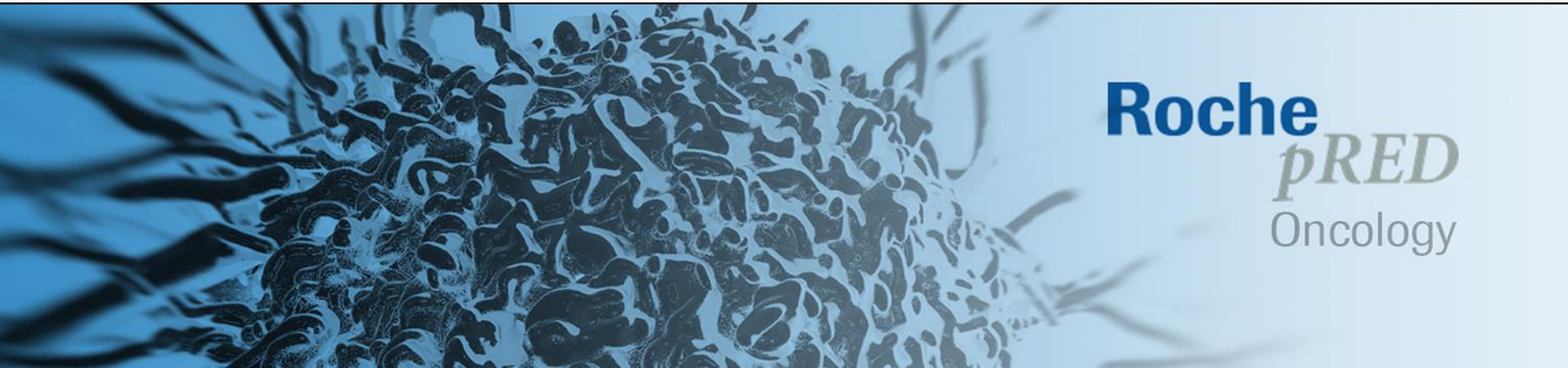
# Cancer immunotherapies: Which efficacy endpoints and statistical analyses to use?

*Daniel Sabanés Bové*

**Basel Biometric Section Spring Seminar, Basel, 28th April 2016**



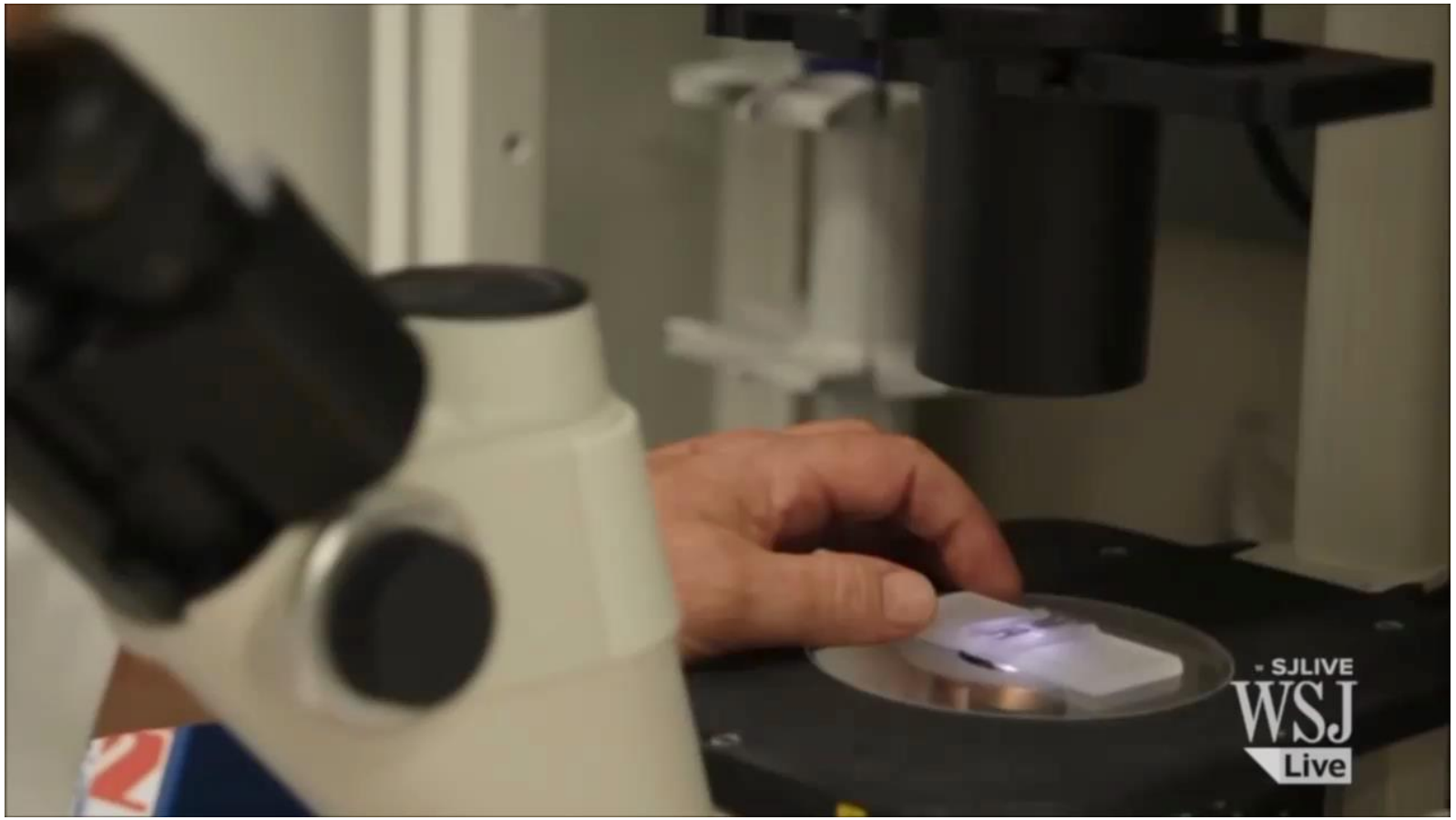
# Cancer immunotherapies

## *Science breakthrough of the year 2013*


"Antibodies (pink) zoom toward a T cell (gray, with CTLA-4 receptor proteins shown in light blue), giving the T cell a push to attack tumor cells. In 2013, new therapies targeting the immune system to treat cancer surged ahead, with promising but still preliminary results in people with many forms of the disease."

# Cancer immunotherapies

## *The ipilimumab (anti-CTLA-4) story*




Full video at https://www.youtube.com/watch?v=ySG2AwpSZmw

See also http://crl.berkeley.edu/discoveries/the-story-of-yervoy-ipilimumab/

# Cancer immunotherapies
## *What makes them different?*

- American Cancer Society: “Cancer Immunotherapy (CIT) is a treatment that uses certain parts of a person’s immune system to fight cancer”

- Important modes of action:
  - *Antibody dependent cell-mediated cytotoxity: → targeted therapy*
    Engineered monoclonal antibodies dock on specific cancer proteins to let the immune system attack the cancer cells, e.g. trastuzumab
  - *Checkpoint inhibition*:
    Inhibit checkpoint proteins on immune cells, basically «taking the brakes off» the immune system, e.g. ipilimumab
  - *(Personalized) cancer vaccines*: Extracted immune cells are exposed to cancer antigens and inserted back to the patient, e.g. Sipuleucel-T
  - *Bispecific T-cell engagement*: One part binds to cancer cells, one to T-cells in order to kill the cancer cell, e.g. blinatumomab

# Cancer immunotherapies

## *Recent FDA approvals of checkpoint inhibitors*

| Date | INN | Approved treatment |
|---|---|---|
| 25/03/2011 | **Ipilimumab** (anti-CTLA-4 mAb) | Unresectable or metastatic melanoma |
| 28/10/2015 | | Stage 3 adjuvant melanoma |
| 22/12/2014 | **Nivolumab** (anti-PD-1 mAb) | Unresectable or metastatic melanoma after ipilimumab |
| 04/03/2015 | | Advanced (metastatic) squamous non-small cell lung cancer (NSCLC) with progression on or after platinum-based chemotherapy |
| 01/10/2015 | | Combination with **Ipilimumab** approved in metastatic melanoma |
| 09/10/2015 | | Approved in addition the non-squamous subtype of NSCLC |
| 23/11/2015 | | Advanced (metastatic) renal cell carcinoma after anti-angiogenic therapy |
| 04/09/2014 | **Pembrolizumab** (anti-PD-1 mAb) | Advanced or unresectable melanoma after ipilimumab |
| 02/10/2015 | | Advanced (metastatic) non-small cell lung cancer (NSCLC) whose disease has progressed after other treatments and with tumors that express a protein called PD-L1 |

# Drug development challenges for CIT
## *Outline of this talk*

Some patients' responses do not follow classical patterns:

- First, lesions grow or new lesions occur, then later they shrink
- Early phase trials must not only look at (classical) responses

Simulations to study statistical impact of CIT mode of action:

- Delayed treatment effect leading to delayed hazard ratio < 1
- Good long-term efficacy and cure rates

Alternative analysis methods:

- Weighted log-rank test (brief)
- Milestone survival

# Drug development challenges for CIT

## *New response patterns*

Some patients' responses do not follow classical patterns:

- First, lesions grow or new lesions occur, then later they shrink
- Early phase trials must not only look at (classical) responses

Simulations to study statistical impact of CIT mode of action:

- Delayed treatment effect leading to delayed hazard ratio < 1
- Good long-term efficacy and cure rates

Alternative analysis methods:

- Weighted log-rank test
- Milestone survival

# RECIST criteria vs. overall survival
## *Ph2 experience checkpoint inhibitor ipilimumab*

- "Four-year survival rates for patients with metastatic melanoma who received ipilimumab in phase II clinical trials*" (Wolchok et al., 2013)

**Table 4.** Best overall response (BOR) in patients with OS >4 years

| BOR by investigator assessment | >4 years OS *n* (%) | | |
|---|---|---|---|
| | CA184-007 (*n* = 28) | CA184-008 (*n* = 27) | CA184-022 (*n* = 33) |
| CR | 5 (17.9) | 2 (7.4) | – |
| PR | 11 (39.3) | 8 (29.6) | 9 (27.3) |
| SD | 7 (25) | 13 (48.1) | 9 (27.3) |
| PD | 4 (14.3) | 3 (11.1) | 15 (45.5) |

BOR, best overall response; CR, complete response; PD, progressive disease; PR, partial response; SD, stable disease.

* "included those who progressed in the parent studies and were retreated in study CA184-025, those who received maintenance therapy and were subsequently retreated upon disease progression in study CA184-025, and those who did not receive further ipilimumab treatment"

- Could indicate that RECIST is not the best measure of benefit for patients treated with immune checkpoint inhibitors

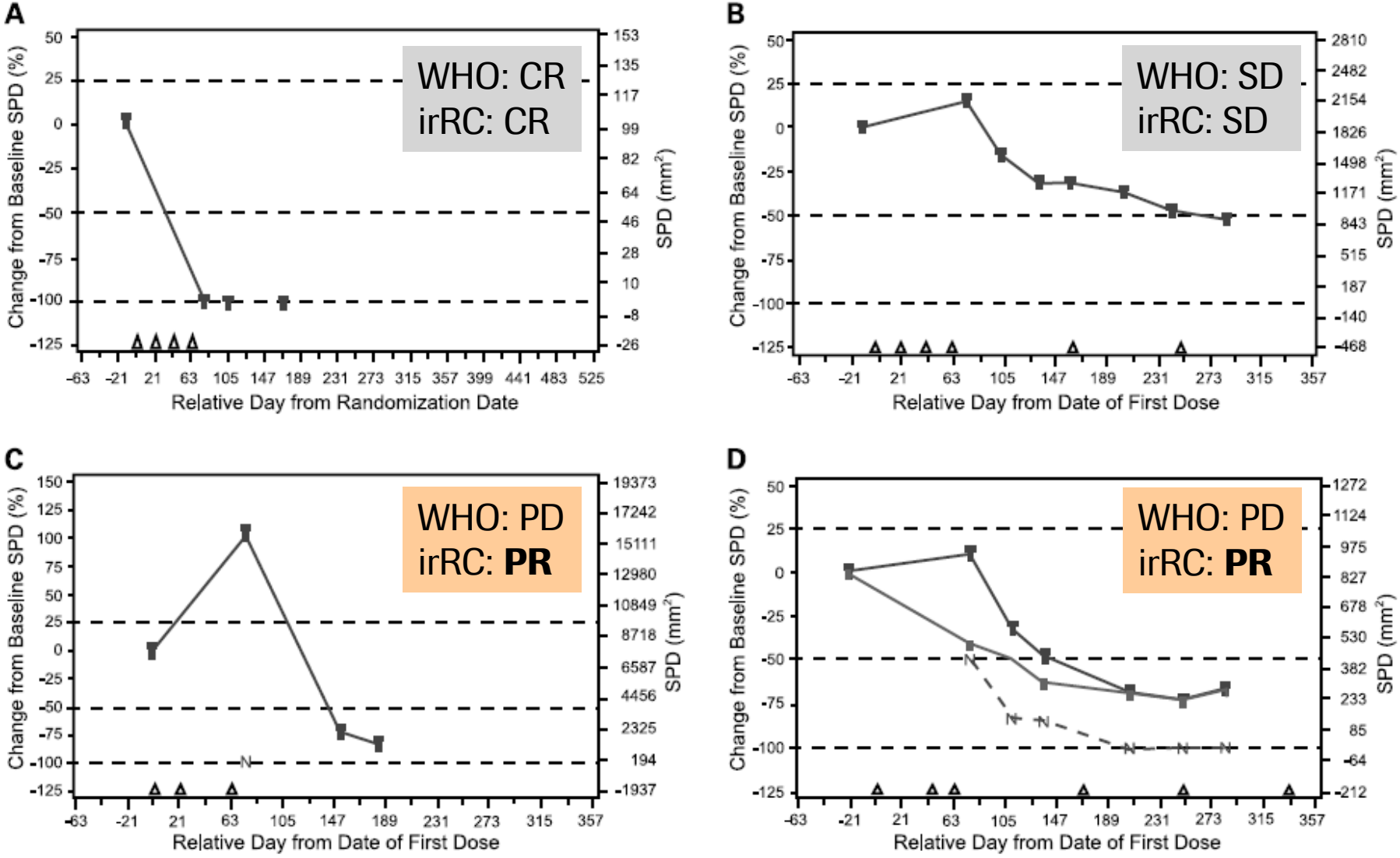
# Concept of "pseudoprogression"
## *Potential reason for "PD yet OS benefit"*

- Observation: New patterns of CT-imaged response to ipilimumab in advanced melanoma clinical trials (Similar later with other CITs)
- Clinical concept: Patterns are new because they…
  - Fulfill standard radiographic criteria for progressive disease (PD)
  - Yet occur in patients who derive clinical/OS benefit from treatment. (prolonged OS compared to patients with PD by both conventional *and* immune-related criteria)
- Biological explanation:
  - Delayed anti-cancer immune activity
  - Transient immune-cell infiltration and inflammation

(Wolchok CCR 2009, Hodi ASCO 2014, Wolchok ASCO 2015)

# Response patterns observed with ipilimumab

## → *Immune-related response criteria (irRC)*




Y-axis: SPD = Sum of products of longest perpendicular diameters

Wolchok et al., 2009

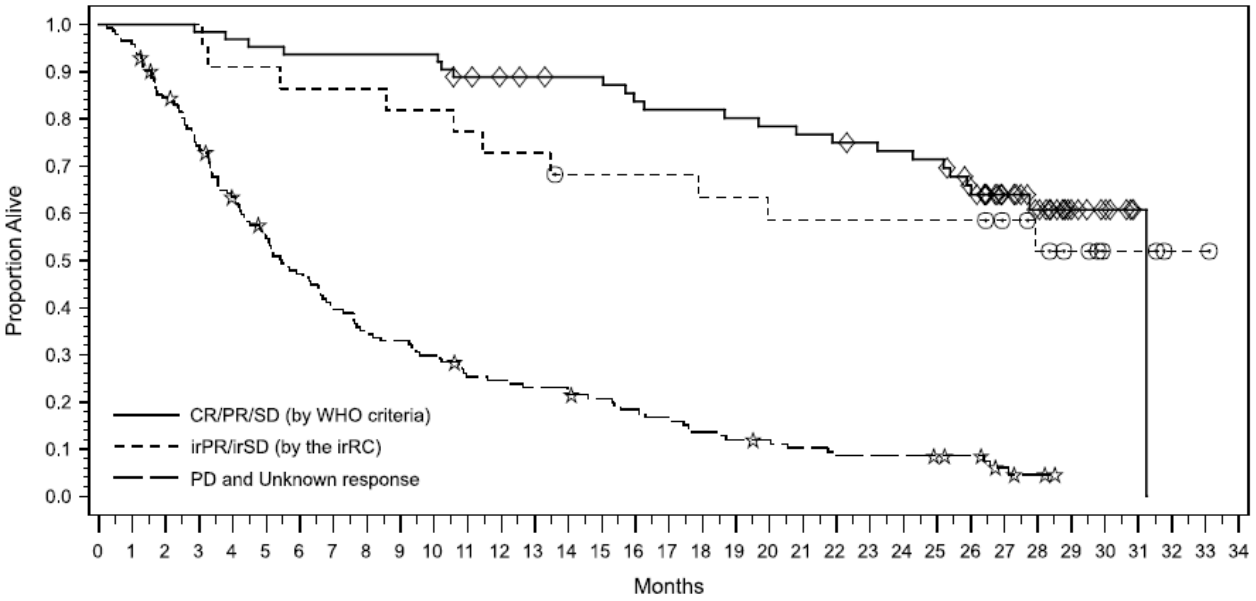
# Overview of different response criteria

## *Classical vs. CIT criteria*

| | WHO | RECIST 1.1 | irRC (Wolchok et al., 2009) | imRECIST (Roche and others) |
|---|---|---|---|---|
| **Tumor burden** | Bidimensional<br>Sum of product of longest perpendicular diameters (SPD) of lesions ≥ 5 x 5 mm | Unidimensional<br>Sum of diameters of lesions ≥ 10 mm | Bidimensional per WHO | Unidimensional per RECIST |
| **New lesions** | Always represent PD | | • **New lesions do not categorically define PD**<br>• **Measurable new lesions are incorporated into the total tumor burden.**<br>• Unmeasurable new lesions preclude CR | |
| **Non-target lesions** | Can contribute to defining CR, PR, SD, and PD | Can contribute to defining CR or PD (unequivocal progression) | Can only contribute to defining CR (complete disappearance required) | |
| **PD** | • ≥ 25% increase in the SPD (WHO) or ≥ 20% increase in the sum of diameters (RECIST) compared with nadir and/or<br>• Unequivocal progression of non-target lesions and/or<br>• Appearance of new lesions<br>• Confirmation of PD not required | | • **Determined only on the basis of measurable disease**<br>• **Negated by subsequent non-PD assessment ≥ 4 weeks from the date first documented (lack of confirmation)** | |

# Ipilimumab melanoma Phase 2 trials
## *WHO/irRC categories vs. OS*

- Kaplan-Meier estimates of overall survival on the basis of best overall response per WHO and irRC for all patients in ipilimumab arms in the CA184-008 and CA184-022 studies (n = 227):

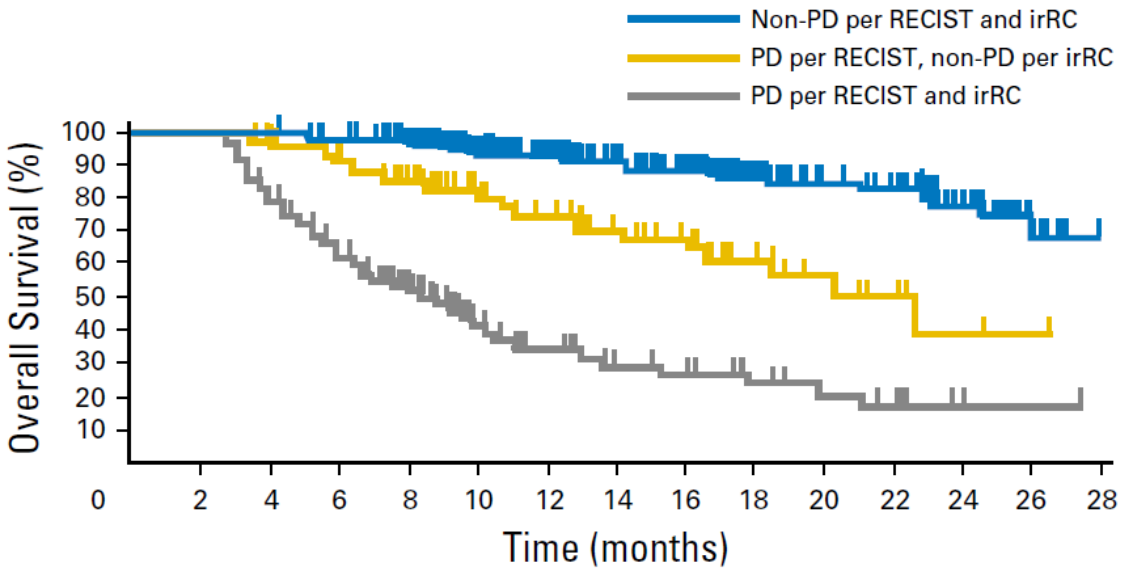



- PD by both WHO and irRC: 142 patients, perform worse than the patients with…
- …PD by WHO but PR/SD by irRC: 22 patients
  ➔ suggests that WHO criteria might underestimate the benefit of ipilimumab in these approximately 10% of patients

Wolchok et al., 2009

# Pembrolizumab melanoma KEYNOTE-001 trial
## *RECIST/irRC categories vs. OS*

- 7% of evaluable patients experienced early or delayed tumor pseudoprogression
- Kaplan–Meier estimates of overall survival on the basis of best overall response per RECIST v1.1 and irRC in patients who survived ≥12 weeks (n = 592) – caveat!




- PD by both RECIST and irRC: 177 patients, perform worse than the patients with …
- …PD by RECIST, but PR/SD by irRC: 84 patients
  ➔ suggests that RECIST v1.1 might underestimate the benefit of pembrolizumab in approximately these 14% of patients

Hodi et al., 2016

# Implications for CIT development
## *Classical response/PFS will not suffice*

- The utility of traditional radiographic response criteria for CIT is limited by the non-classical tumor kinetics ("pseudoprogression") observed in some patients with clinical benefit
- Response criteria especially relevant in early phase trials (phase 1/2)
  - It will not be enough to only look at PR/CR by RECIST
  - Prevalence of pseudoprogression and atypical response patterns likely to differ between different mode of actions (MoA, e.g. anti-CTLA-4 vs. anti-PD-1/PD-L1) and different indications (e.g. melanoma vs. NSCLC)
- irPFS can be defined based on the irRC PD timepoint (for phase 2/3)
- Standardization and validation across multiple trials still to be achieved:
  *«As a community, we must advocate the sharing of clinical data from multiple studies and immunotherapy agents to greatly hasten and provide rigor to this effort»*
  (Hodi et al., 2016)

# Drug development challenges for CIT

## *Delayed separation and cure rates impact*

Some patients' responses do not follow classical patterns:

- First, lesions grow or new lesions occur, then later they shrink
- Early phase trials must not only look at (classical) responses

Simulations to study statistical impact of CIT mode of action:

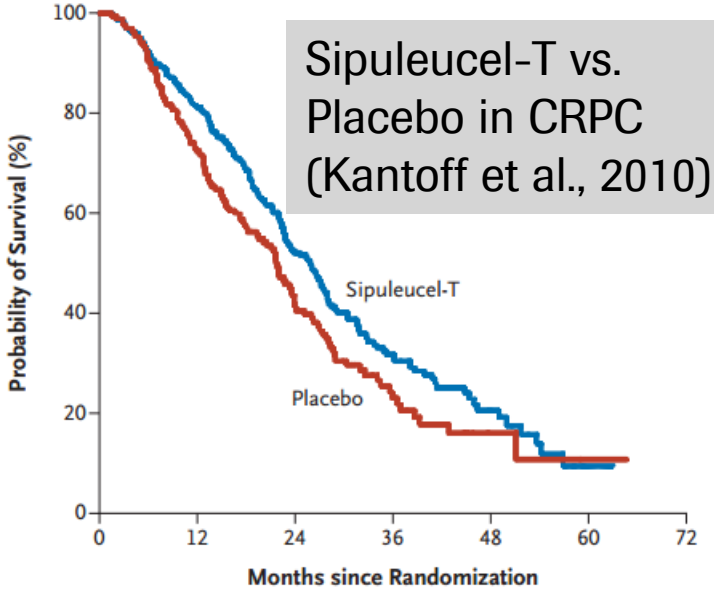
- Delayed treatment effect leading to delayed hazard ratio < 1
- Cure rates > 0

Alternative analysis methods:

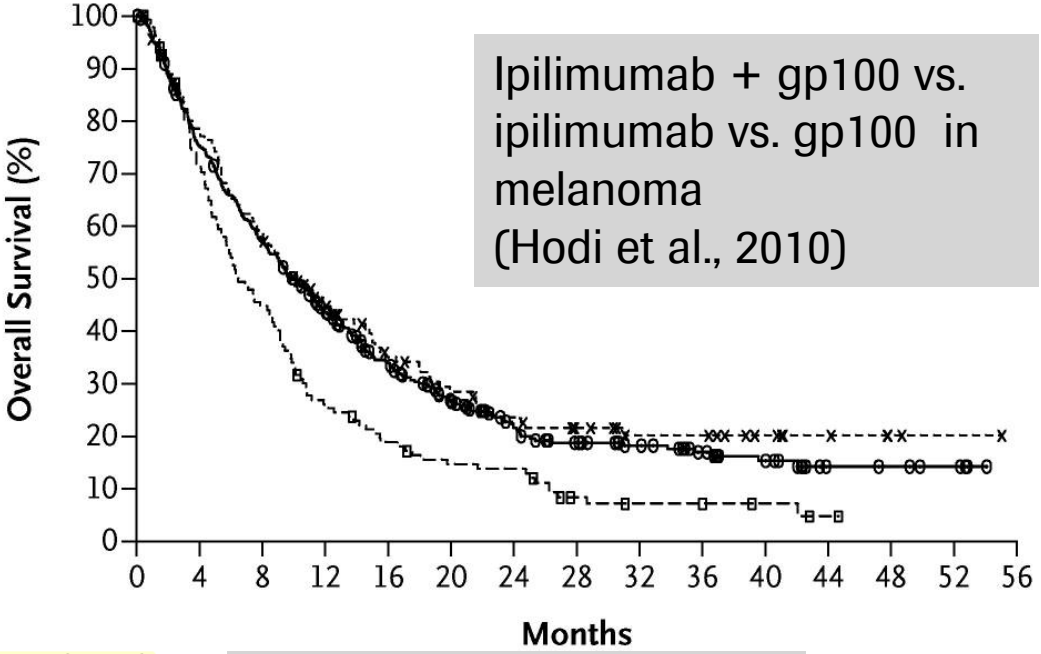
- Weighted log-rank test
- Milestone survival

# Motivation of simulation study

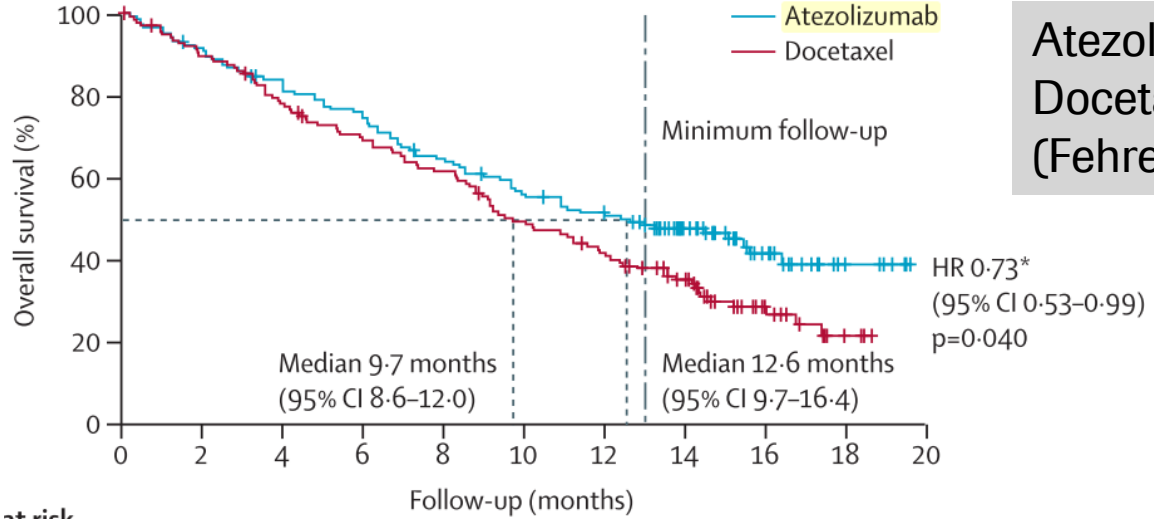
## *Observed delayed separation and cure rates*




Sipuleucel-T vs. Placebo in CRPC (Kantoff et al., 2010)




Ipilimumab + gp100 vs. ipilimumab vs. gp100 in melanoma (Hodi et al., 2010)



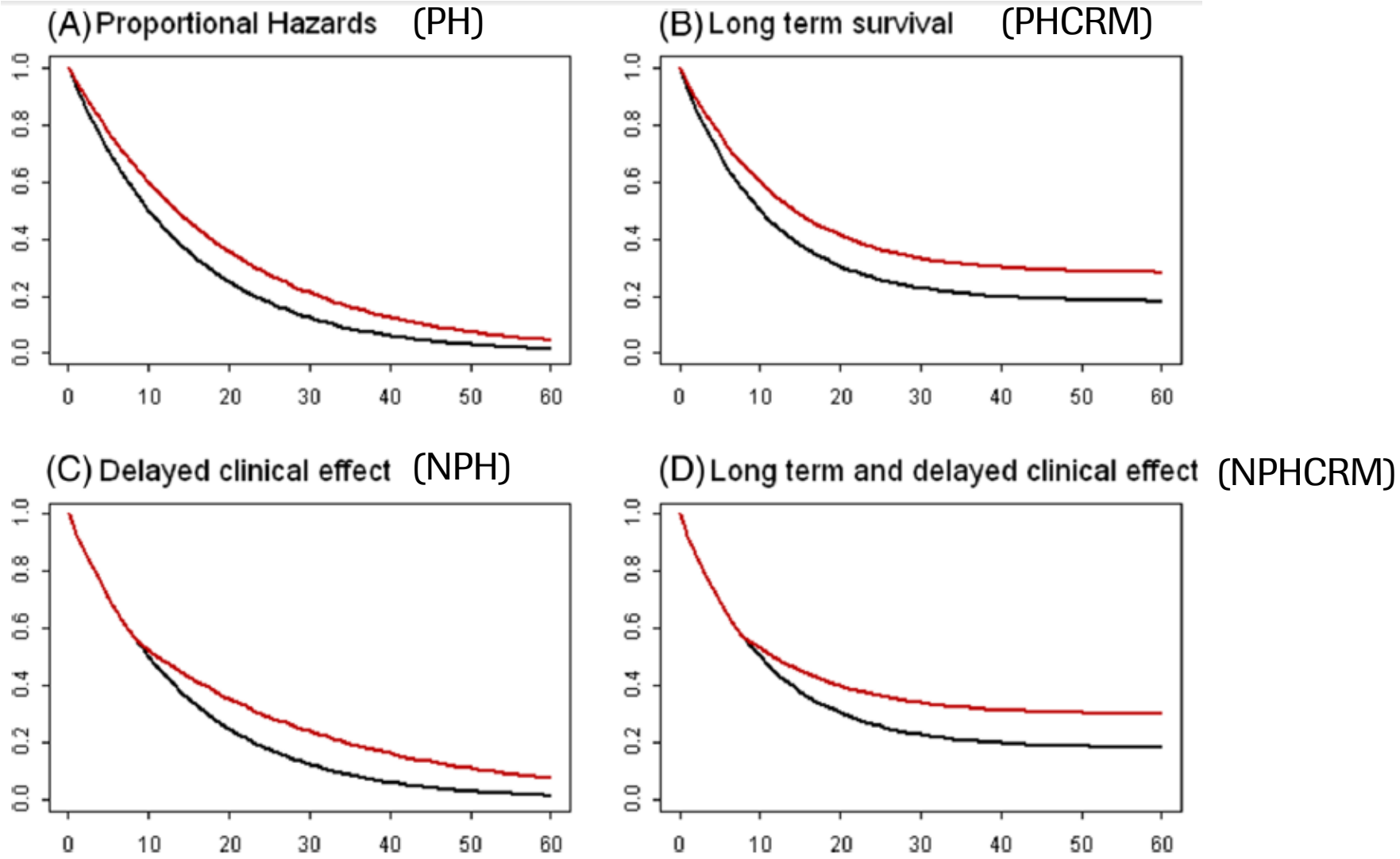

Atezolizumab vs. Docetaxel in NSCLC (Fehrenbacher et al., 2016)

# Simulation study to examine impact on studies
## *Setup (similar to Chen, 2013)*

- Design and analysis using proportional hazards assumption and log-rank test to have 90% power to detect HR = 0.75 at a type I error rate of 5%
- Hence, 508 events are required. Median OS in control arm assumed to be 12.5 months. In order to recruit 680 patients, 34 months accrual time is projected.
- Examine using simulations (with R-package TTESimu by Carrie Li):
  - Cure rates: Number of cured patients drawn from binomial distributions. The HR only applies to remaining non-cured patients.
  - Non-proportional hazards with a delayed separation of the two treatment arms (i.e, piecewise exponential distribution with HR = 1 until time of separation and HR < 1 afterwards)
  - Combination of both

# Simulation study

## *Illustration of setup (Chen, 2013)*



# Simulation study
## *Results*

| Title | Cure Rates | HR | Delay | Power | Type I error | Length under H1 | Length under H0 |
|---|---|---|---|---|---|---|---|
| Proportional Hazards (PH) | 0% | 0.75 | 0 | 0.91 | 0.05 | 48 | 44 |
| PH with Cure Rates (PHCRM) | 10% / 18% | 0.85 | 0 | 0.93 | 0.055 | **59** | **52** |
| Non-PH (NPH) | 0% | 1 / 0.75 | 3 | **0.74** | 0.05 | 47.5 | 44 |
| Non-PH with Cure Rates (NPHCRM) | 10% / 18% | 1 / 0.85 | 3 | **0.88** | 0.055 | **59** | **52** |

Based on 2000 simulated trials for each H0/H1 scenario

# Simulation study
## *Interpretation*

- Cure Rates …
  - prolong the accumulation of the required number of events, therefore the trial can be substantially longer (here: ca.+20%)
  - If the HR in the non-cured patients was the same as the assumed overall HR, then the trial would be over-powered (not shown in table) (→ that is why we used higher HR = 0.85 in non-cured for simulations)
- Delayed separation …
  - Without cure rate (NPH) leads to lower power to detect the effect, because the overall estimated HR will be higher
  - On the other hand, if there are relevant cure rates (NPHCRM) the power (difference to Chen results!) and duration can be similar to PHCRM.
- Results from a Ph2 trial could be used to detail simulation setup for Ph3.

# Drug development challenges for CIT

## *Alternative analysis methods*

Some patients' responses do not follow classical patterns:

- First, lesions grow or new lesions occur, then later they shrink
- Early phase trials must not only look at (classical) responses

Simulations to study statistical impact of CIT mode of action:

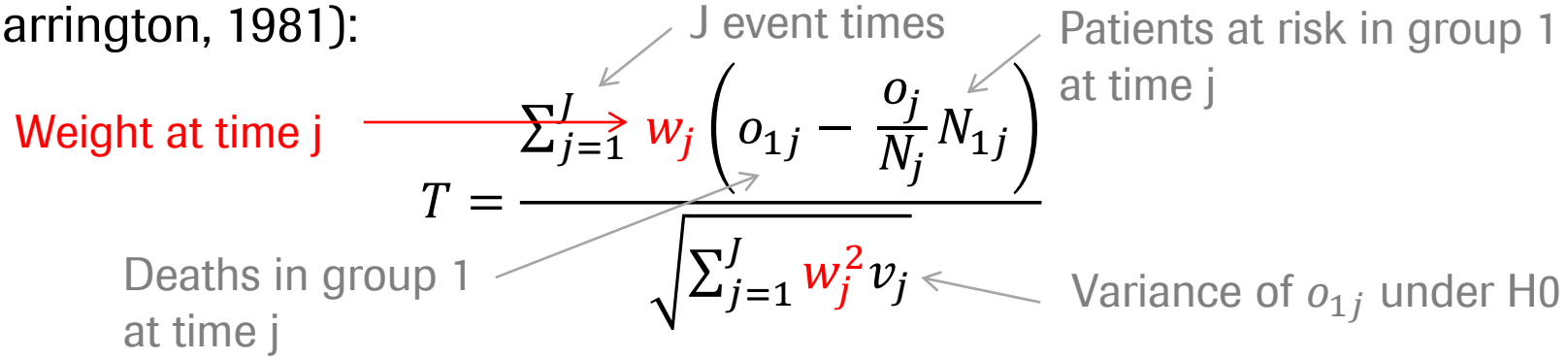
- Delayed treatment effect leading to delayed hazard ratio < 1
- Cure rates > 0

Alternative analysis methods:

- Weighted log-rank test (brief)
- Milestone survival

# Weighted log-rank test
## *Brief outline of the idea*

- The proportional hazards assumption is violated for delayed separation, hence the log-rank test is not optimal. Alternative is weighted log-rank test
- Idea is to plug weights (depending on the sample size *n*) into the log-rank statistic, motivated by the H1 we would like to detect (Fleming and Harrington, 1981):




- Specific family of weights is $W(t) = \hat{S}(t)^{\rho} \left(1 - \hat{S}(t)\right)^{\gamma}$, Log-rank test is then special case with $\rho = \gamma = 0$.
- Power will depend on adequacy of the specified weight function

# Milestone survival
## *Definition and Motivation (Chen, 2015)*

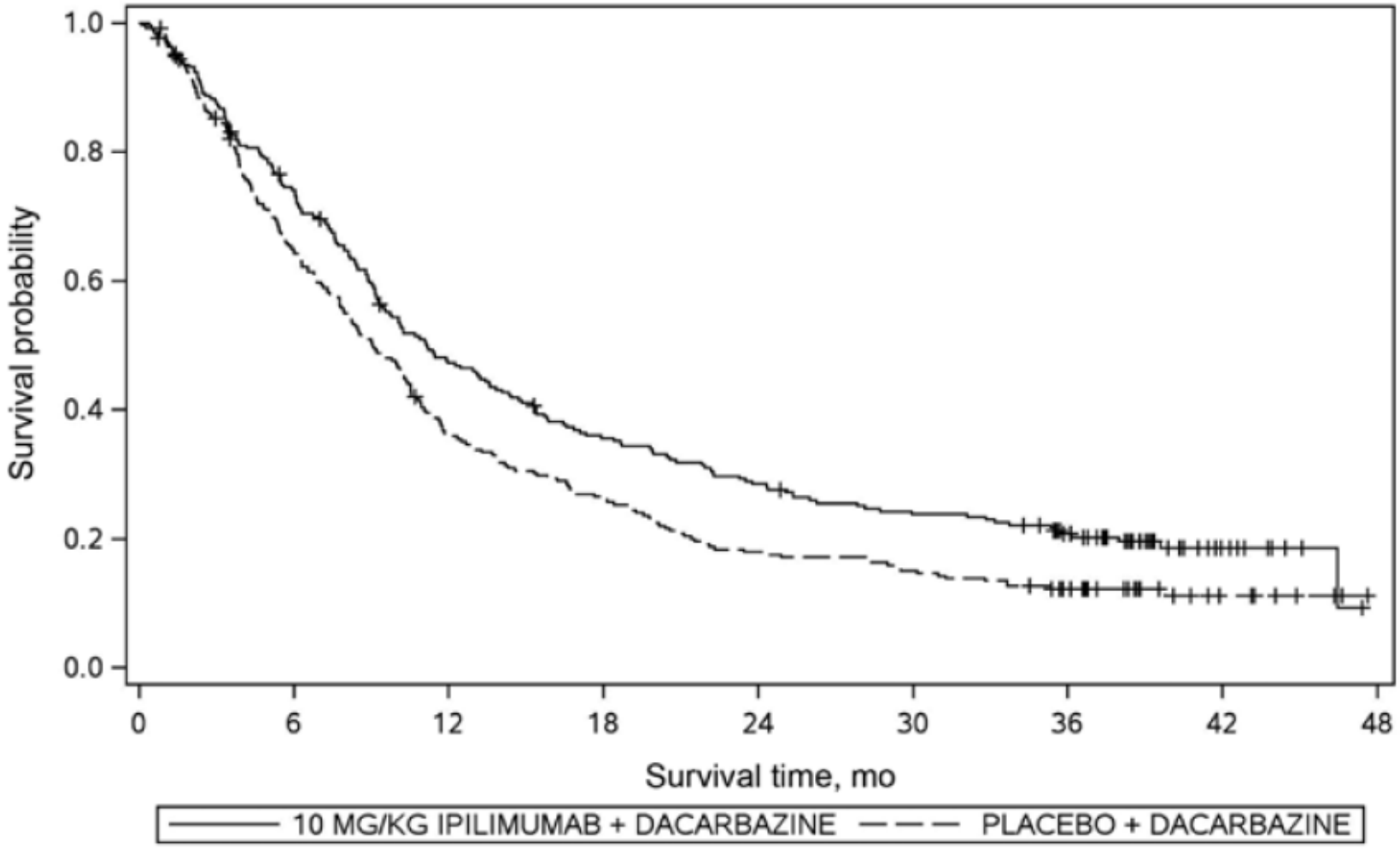
- Milestone survival analysis is a cross-sectional assessment of the OS data at the prespecified time point comparing Kaplan-Meier survival probabilities (see Klein et al., 2007 for comparison techniques)
  - Timepoint often represents a clinically meaningful benchmark, after which the response is deemed stable
  - Proposed to be conducted in the first cohort of randomized patients, rather than in the entire study population, such that all analysed patients have reached milestone time (to ensure robustness)
- Approximately compares cure rates if long-term timepoint is chosen
- Main motivation is gaining time, especially when cure rates / delayed separation of the survival curves is expected
  - Interim analysis time-point can be planned:
    Only depends on enrollment of the first cohort
  - “Certain level” of multiplicity adjustment is warranted

# Milestone survival
## *Example: Ipilimumab Ph3 study*

- Ph3 comparing Ipilimumab vs. Placebo in combination with dacarbazine
- 500 randomized patients (250 per arm) and a total of 416 deaths were needed to provide approximately 90% power to detect a $\mathrm{HR} = 0.727$ (PH assumption!)
- It was estimated that it would take 17 months to complete the enrollment and another 17 months of follow-up, so a total study duration of 34 months
- The study was initiated in August 2006. Turns out that:
  - Final analysis in March 2011! ➔ 56 months total!
  - Still only 414 events (two events less than prespecified)
- Final result HR estimate = 0.72 (95% CI = 0.59 to 0.87, $P < .001$)

# Milestone survival

## *Example: Ipilimumab Ph3 Kaplan-Meier plot*



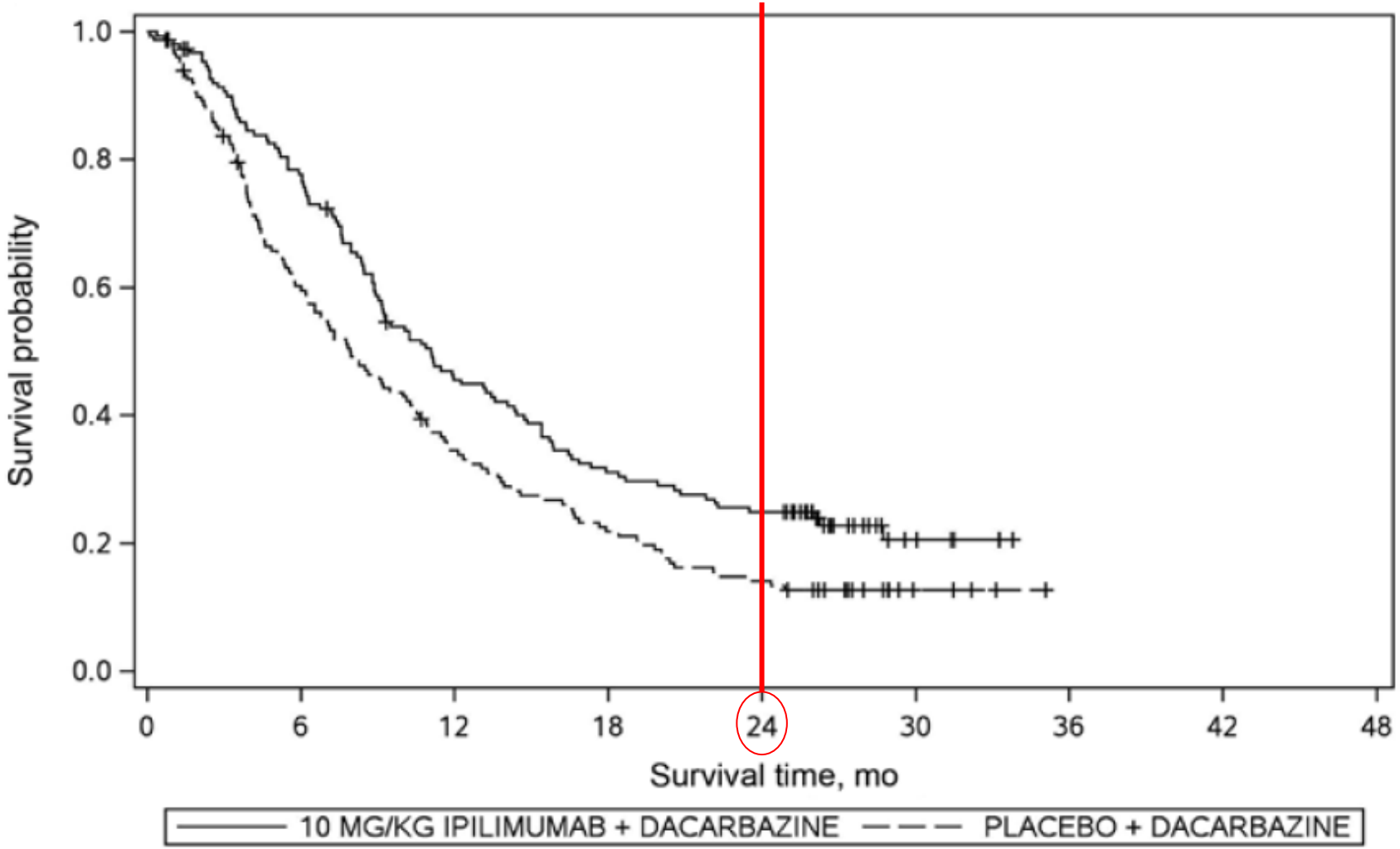
# Milestone survival

## *Example: Ipilimumab Ph3 retrospective design*

- Assume NPHCRM with delay of 4 months until separation, and 10% cure rate in control arm ➔ simulation recovers observed trial duration
- Interim analysis using 2 year milestone survival on the first 300 randomized patients.
- Ensure overall type I error of $\alpha = 0.05$ by:
  - Nominal $\alpha = 0.025$ at interim milestone survival analysis
  - Nominal $\alpha = 0.0328$ for final log-rank test on all patients (not immediately obvious for me how this is calculated…)
- In this case the interim analysis gave a nominal p-value of 0.021
  - Could have potentially saved 18 months until efficacy was declared

# Milestone survival

## *Example: Ipilimumab Ph3 interim analysis*



# Conclusion

## *Active engagement of statisticians needed in CIT*

- Non-standard behavior of CITs leads to non-standard trial results.
- Hence, just using the standard designs and analyses can potentially lead to underpowered or longer than expected trials.
  - Could miss a working molecule in early phase because of not increased objective response (CR/PR) rate or atypical responses
  - Could declare futility in a pivotal study because the separation of survival curves occurs late
  - Further risk from too aggressive futility interim analyses (not shown)
- Therefore, active engagement of statisticians is more than ever needed in designing efficient and successful CIT trials
  - Use of existing information about the molecule / MoA, Simulations to assess impact of delayed separation / cure rates
  - Cross-company and academia collaborations are necessary

*Thank you! Questions?*

# References
## *Journal articles*


- Chen, T.-T. (2013). Statistical issues and challenges in immuno-oncology. *Journal for ImmunoTherapy of Cancer*, *1*(1), 18. doi:10.1186/2051-1426-1-18
- Chen, T.-T. (2015). Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors. *Journal of the National Cancer Institute*, *107*(9), djv156. doi:10.1093/jnci/djv156
- Fehrenbacher, L., Spira, A., Ballinger, M., Kowanetz, M., Vansteenkiste, J., Mazieres, J., … Rittmeyer, A. (2016). Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer ( POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *The Lancet*, *6736*(16), 1–10. doi:http://dx.doi.org/10.1016/S0140-6736(16)00587-0
- Fleming, T. R. & Harrington, D. P. (1981). A class of hypothesis tests for one and two sample censored survival data. *Communication in Statistics Theory and Methods, 10*(8):763-794
- Hodi, F. S., Hwu, W.-J., Kefford, R., Weber, J. S., Daud, A., Hamid, O., … Wolchok, J. D. (2016). Evaluation of Immune-Related Response Criteria and RECIST v1.1 in Patients With Advanced Melanoma Treated With Pembrolizumab. *Journal of Clinical Oncology*. doi:10.1200/JCO.2015.64.0391

# References
## *Journal articles (cont.)*


- Kantoff, P. W., Higano, C. S., Shore, N. D., Berger, E. R., Small, E. J., Penson, D. F., … Schellhammer, P. F. (2010). Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer. *New England Journal of Medicine*, *363*(5), 411–422. doi:10.1056/NEJMoa1001294
- Klein, J. P., Logan, B., Harhoff, M., & Andersen, P. K. (2007). Analyzing survival curves at a fixed point in time. *Statistics in Medicine*, *26*(24), 4505–4519. doi:10.1002/sim.2864
- Kosorok, M. R., Lin, C.-Y. (1999) The Versatility of Function-Indexed Weighted Log-Rank Statistics. *Journal of the American Statistical Association, 445*(94)*,* 320-332.
- Wolchok, J. D., Hoos, A., O'Day, S., Weber, J. S., Hamid, O., Lebbé, C., … Hodi, F. S. (2009). Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *15*(23), 7412–20. doi:10.1158/1078-0432.CCR-09-1624
- Wolchok, J. D., Weber, J. S., Maio, M., Neyns, B., Harmankaya, K., Chin, K., … Lebbé, C. (2013). Four-year survival rates for patients with metastatic melanoma who received ipilimumab in phase II clinical trials. *Annals of Oncology*, *24*(8), 2174–2180. doi:10.1093/annonc/mdt161

*Doing now what patients need next*