Design for immuno-oncology clinical trials with non-proportional hazards patterns

Zhenzhen Xu

U.S. Food and Drug Administration

Statistical Issues in Clinical Trials 2023

Joint work with

Bin Zhu Ph.D.

National Cancer Institute
Principle Investigator

Yongsoek Park Ph.D.

University of Pittsburg
Assistant Professor

Boguang Zhen Ph.D.

FDA
Branch Chief

Unkyung Lee Ph.D.

FDA
Postdoc Fellow

Acknowledgement

The work is supported by the Chief Scientist Intramural Challenge Grant of the U.S. Food and Drug Admiration

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's view or policies

Introduction

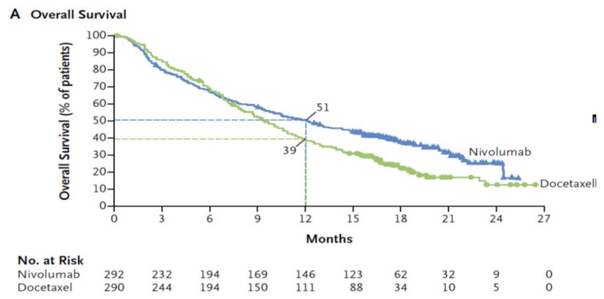
Challenges in immuno-oncology (IO) trials

- Unprecedented growth outstripped development of design and analysis
- Non-proportional hazards (NPH) patterns manifested in Kaplan-Meier curves

NPH Patterns

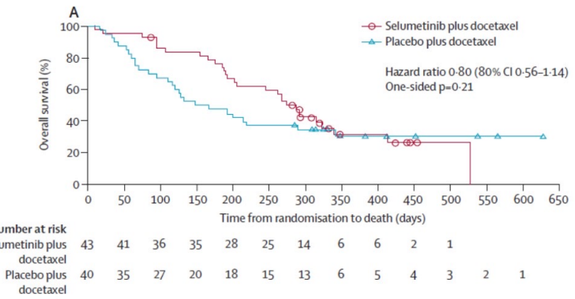
Delayed effect

Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer



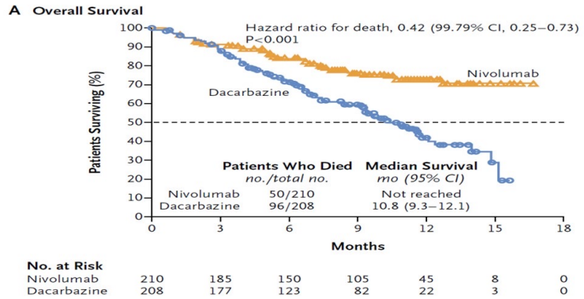
Diminishing effect

Selumetinib plus docetaxel for *KRAS*-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study



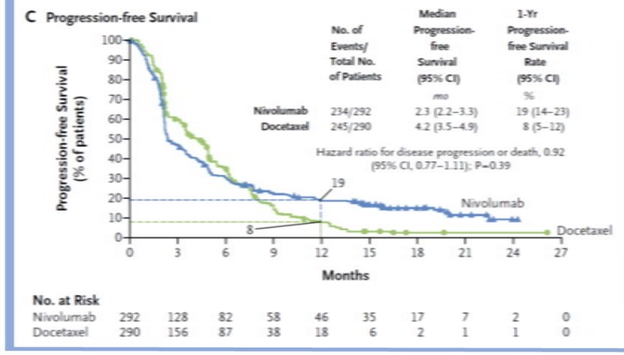
Crossing hazards

Nivolumab in Previously Untreated Melanoma without *BRAF* Mutation

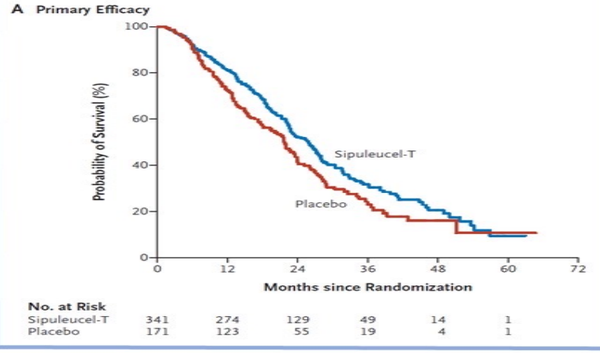


Combination

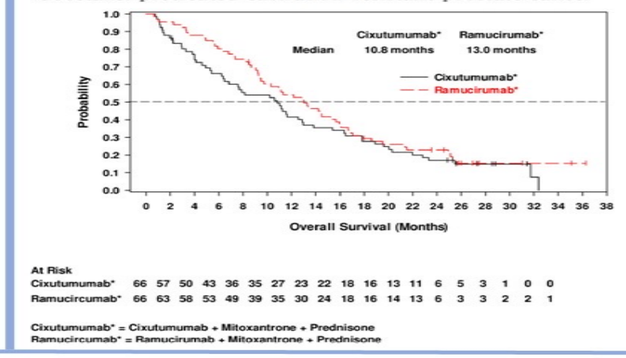
Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer



Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer



A randomised non-comparative phase II trial of cixutumumab (IMC-A12) or ramucirumab (IMC-1121B) plus mitoxantrone and prednisone in men with metastatic docetaxel-pretreated castration-resistant prostate cancer



Statistical Challenges of NPH issue:

- Violate proportional hazards assumption
- Cause underpowered or even falsely negative studies

Question of Interest

- How to design adequate and well-controlled IO trials?
- How to mitigate the occurrence of complex NPH patterns?

Our strategy

- **Cause**: What are underlying cause or causes behind NPH patterns?
- **Solution**: Targeting causes, develop proper design and analysis strategies

Outline of the talk

- Delayed Effect Pattern
  - Cause: Indirect working mechanism
  - Solution: APPLE, APPLE+

- NPH Patterns
  - Causes: mechanism + heterogeneity
  - Solution: PRIME, PRIME+

Delayed Effect Pattern

Causes of Delayed Effect Pattern

- Primary causes: Indirect mechanism of action
  - Frontline Investigation of Revlimid and Dexamethasone vs Standard Thalidomide (FIRST) study
    - Revlimid: Immunomodulatory drug
    - Transplant-ineligible patients with Myeloma

Motivating example

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

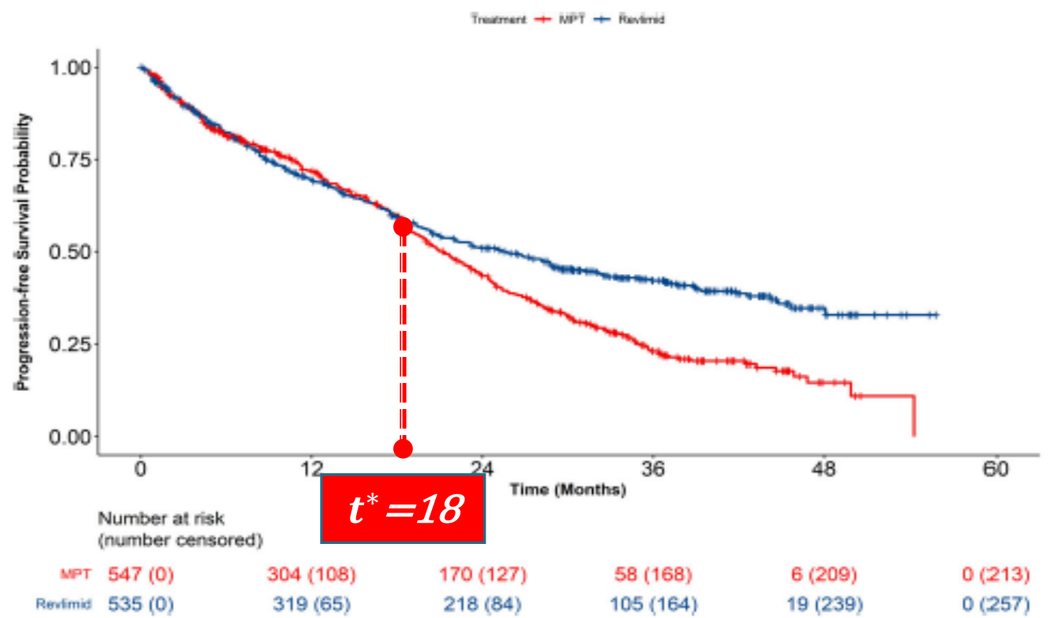


Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Motivating example

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

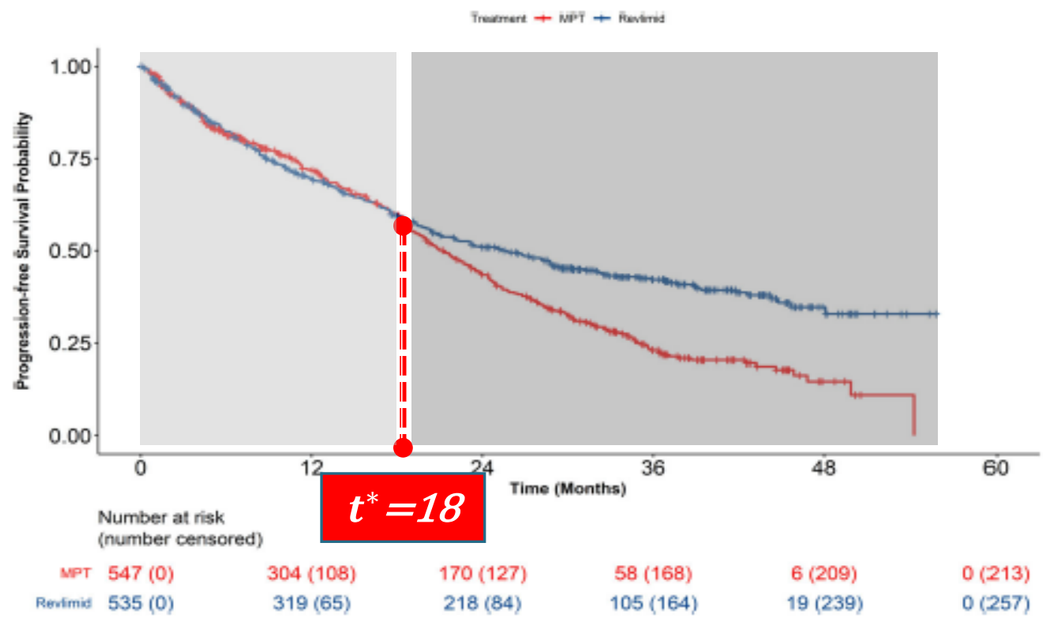


Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Motivating example

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

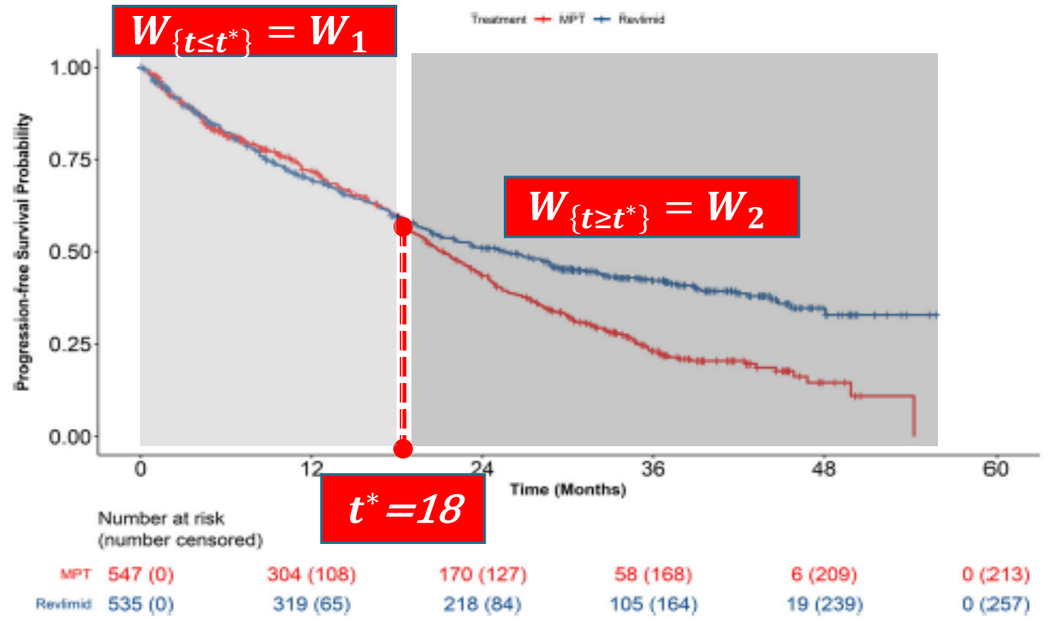


Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Motivating example

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

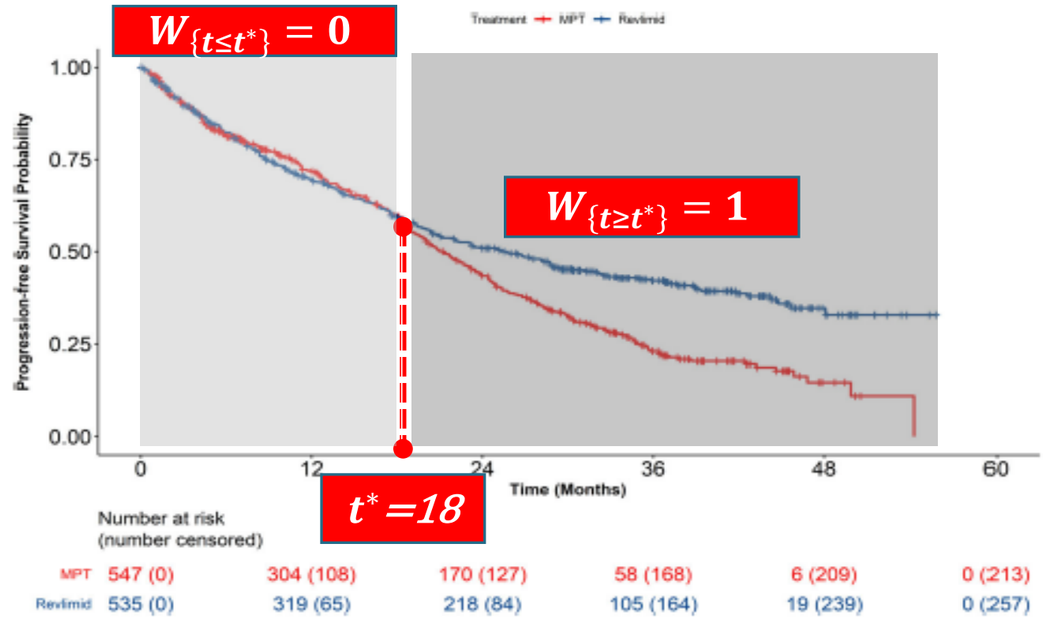


Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Piecewise Weighted Logrank Test

Theorem 1. Under fixed delayed scenario, the optimal weights $W_j^* = argmax\{Pow(w_j)\}$ need to satisfy that $\boldsymbol{W_j^* \propto log\{\lambda(t_j)\}}$

$$H_0: \lambda(t) = 1 \quad vs \quad H_1: \lambda(t) = \begin{cases} 1, & t < t^* \\ < 1, & t \geq t^* \end{cases}$$

$$\boldsymbol{W^*(t) = \begin{cases} 0, & t < t^* \\ 1, & t \geq t^* \end{cases}}$$

APPLE & SEPPLE

Piecewise Weighted Logrank Test:

- Analytic Power calculation based on Piecewise-weighted Logrank test (APPLE)
- Simulation-based Empirical Power calculation based on Piecewise-weighted Logrank test (SEPPLE)

Pros and Cons

- **Pros:**
  - Practical applications:
  - FDA Science Board:

FDA Chief Scientist Publication Award:
An exceptional manuscript with immediate impact that may speed availability of cancer therapies

- **Cons:**
  - Fixed Lag Effect scenario: Each subject takes same lag t^* (biologically implausible)
  - t^* can be properly specified in advance (mis-specification risk)

Motivating example

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

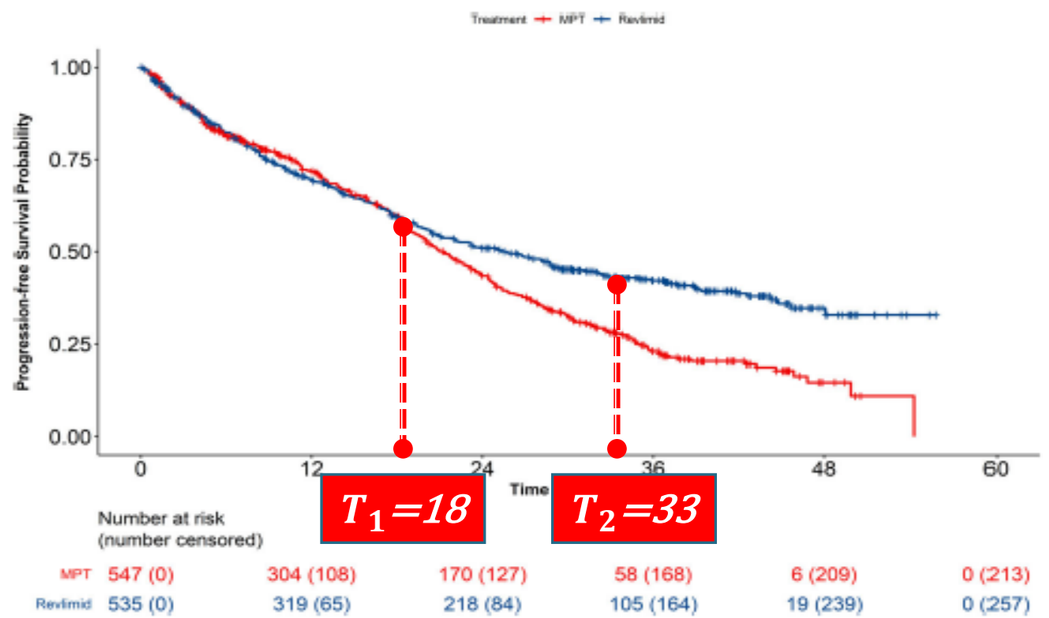
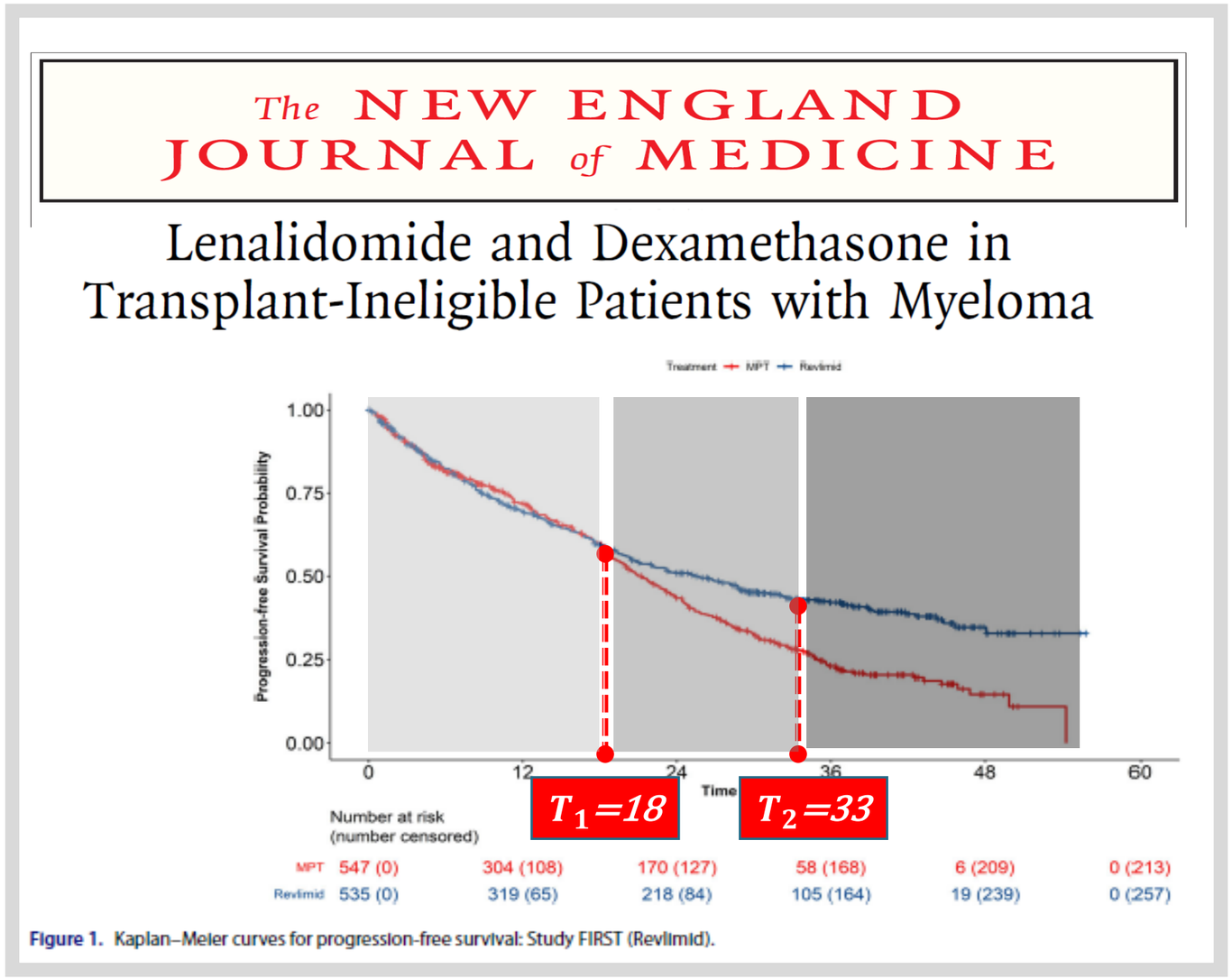


Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Motivating example



The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Assumptions: Random lag effect scenario

Each subject takes a specific lag $t^*_{ind} \sim \text{Dist}(T_1, T_2)$

- T_1: Patient's shortest possible treatment lag time
- T_2: Patient's longest possible treatment lag time

Generalized Piecewise Weighted Logrank Test

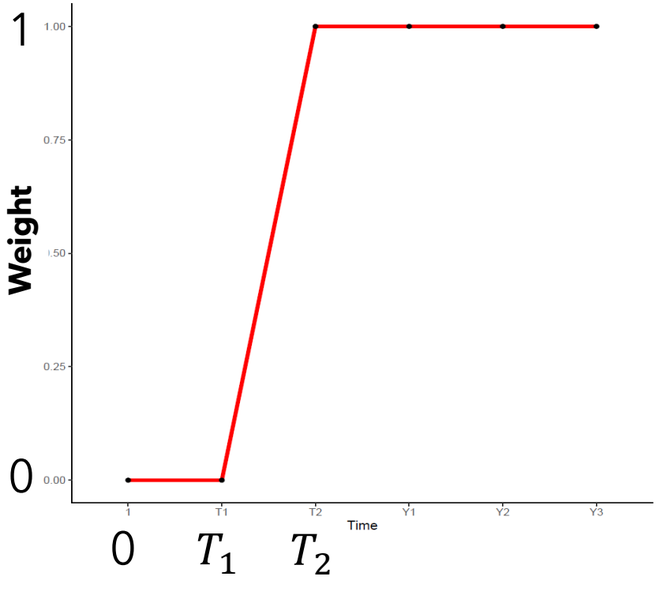
Theorem 2. Under random delayed scenario, the optimal weights $W_j^* = argmax\{Pow(w_j)\}$ need to satisfy that $\boldsymbol{W_j^* \propto F_*(t_j)}$

$$H_0: \lambda(t) = 1 \quad vs \quad H_1: \lambda(t) = f(x) = \begin{cases} 1, & t < T_1 \\ \lambda_2^{\boldsymbol{g(t)}}, & T_1 < t \leq T_2 \\ \lambda_2, & t > T_2 \end{cases}$$

$$\boldsymbol{W^*(t) = F_*(t)}$$

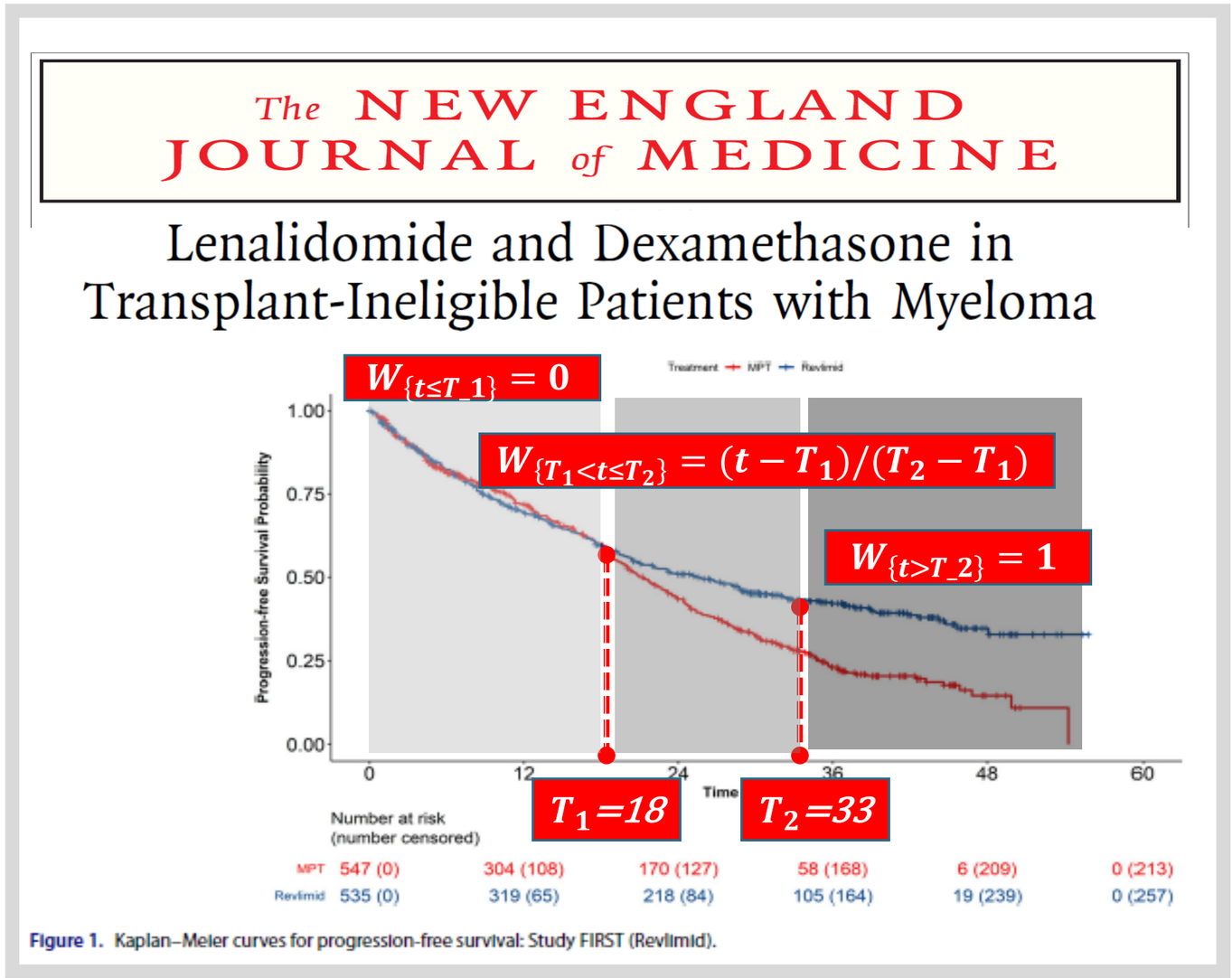
Generalized Piecewise Weighted Logrank Test

If the lag t^*_{ind} follows a uniform distribution on $[T_1, T_2\ ]$:



$$w^*(t) = F_{*u}(t) = \begin{cases} w_1^*(t) = 0, & t \le T_1 \\ w_2^*(t) = (t - T_1)/(T_2 - T_1), & T_1 < t \le T_2 \\ w_3^*(t) = 1, & t > T_2 \end{cases}$$

Motivating example



The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

$W_{\{t \le T\_1\}} = 0$

$W_{\{T_1 < t \le T_2\}} = (t - T_1)/(T_2 - T_1)$

$W_{\{t > T\_2\}} = 1$

$T_1 = 18$

$T_2 = 33$

Treatment MPT Revlimid

Progression-free Survival Probability

1.00 0.75 0.50 0.25 0.00

0 12 24 36 48 60

Time

Number at risk (number censored)

MPT	547 (0)	304 (108)	170 (127)	58 (168)	6 (209)	0 (213)
Revlimid	535 (0)	319 (65)	218 (84)	105 (164)	19 (239)	0 (257)

Figure 1. Kaplan–Meier curves for progression-free survival: Study FIRST (Revlimid).

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Fixed scenario with $t^* = 6$	
PW-Logrank $t^* = 6$	79%

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Fixed scenario with $t^* = 6$	
PW-Logrank $t^* = 6$	79%
PW-Logrank $\mathbf{t^m = 1}$	**63%**
PW-Logrank $\mathbf{t^m = 11}$	**64%**

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Fixed scenario with $t^* = 6$	
PW-Logrank $t^* = 6$	79%
PW-Logrank $\mathbf{t^m = 1}$	**63%**
PW-Logrank $\mathbf{t^m = 11}$	**64%**
GPW-Logrank $[T_1, T_2]$=[1,11]	76%
GPW-Logrank $[T_1, T_2]$=[1,9]	76%
GPW-Logrank $[T_1, T_2]$=[3,9]	78%

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Random scenario with $[T_1, T_2]$=[3,9]	
GPW-Logrank $[T_1^*, T_2^*]$=[3,9]	80%

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Random scenario with $[T_1, T_2]$=[3,9]	
GPW-Logrank $[T_1^*, T_2^*]$=[3,9]	80%
GPW-Logrank $[T_1^m, T_2^m]$=[1,9]	79%
GPW-Logrank $[T_1^m, T_2^m]$=[3,11]	79%

Advantage of GPW Logrank test vs PW Logrank test

Test	Power
True parameter setting: Random scenario with $[T_1, T_2]$=[3,9]	
GPW-Logrank $[T_1^*, T_2^*]$=[3,9]	80%
GPW-Logrank $[T_1^m, T_2^m]$=[1,9]	79%
GPW-Logrank $[T_1^m, T_2^m]$=[3,11]	79%
PW-Logrank t^m = 1	**66%**
PW-Logrank t^m = 11	**69%**

APPLE+, SEPPLE+

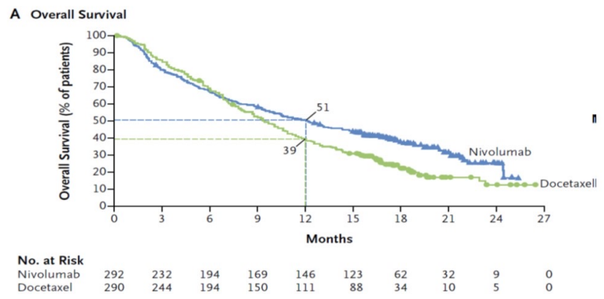
Generalized Piecewise Weighted Logrank Test

- APPLE ⟹ APPLE+
- SEPPLE ⟹ SEPPLE+

How to deal with general NPH Patterns?

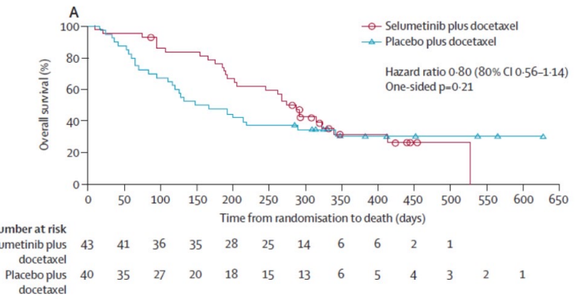
Delayed effect

Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer



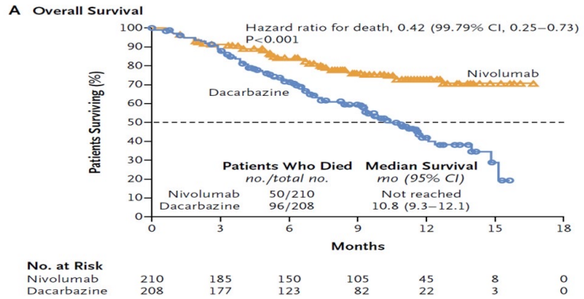
Diminishing effect

Selumetinib plus docetaxel for *KRAS*-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study



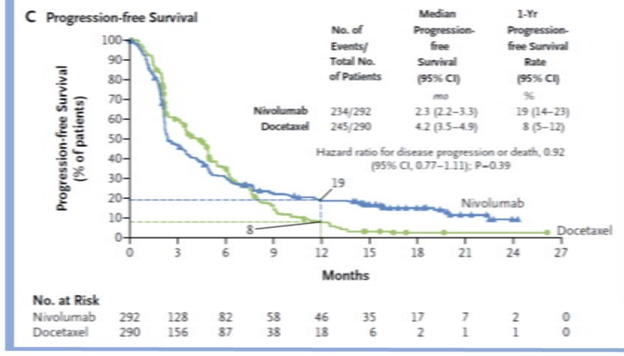
Crossing hazards

Nivolumab in Previously Untreated Melanoma without *BRAF* Mutation

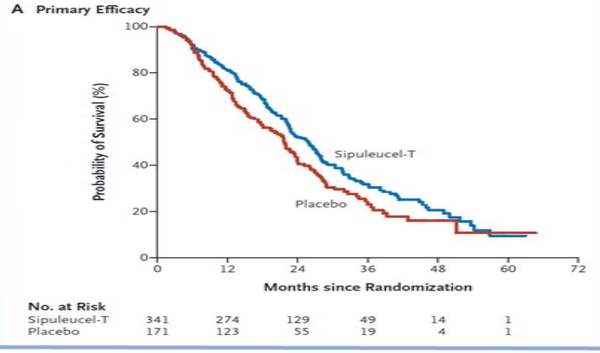


Combination

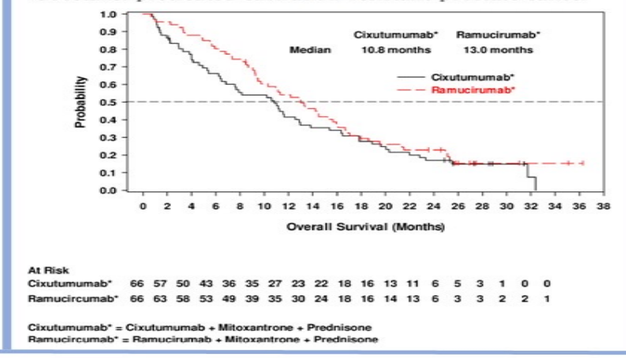
Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer



Sipuleucel-T Immunotherapy for Castration-Resistant Prostate Cancer



A randomised non-comparative phase II trial of cixutumumab (IMC-A12) or ramucirumab (IMC-1121B) plus mitoxantrone and prednisone in men with metastatic docetaxel-pretreated castration-resistant prostate cancer



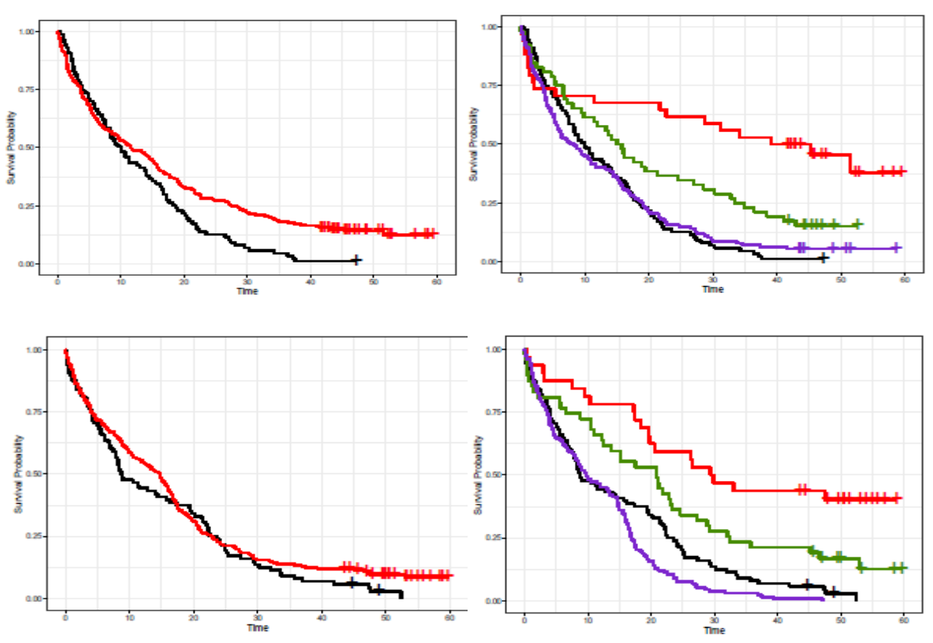
NPH Patterns

Causes of NPH Patterns

- Possible causes: Indirect mechanism of action
  - What are underlying causes behind other NPH patterns?
- There may be more than a working mechanism…

Elephant In The Room

- A limited percentage of treated subjects respond whereas others don't
  - Are we treating heterogeneous patients $\Longrightarrow$ NPH?



A real study

- A limited percentage of treated subjects respond whereas others don't
  - Are we treating heterogeneous patients $\Longrightarrow$ NPH?

The NEW ENGLAND JOURNAL of MEDICINE

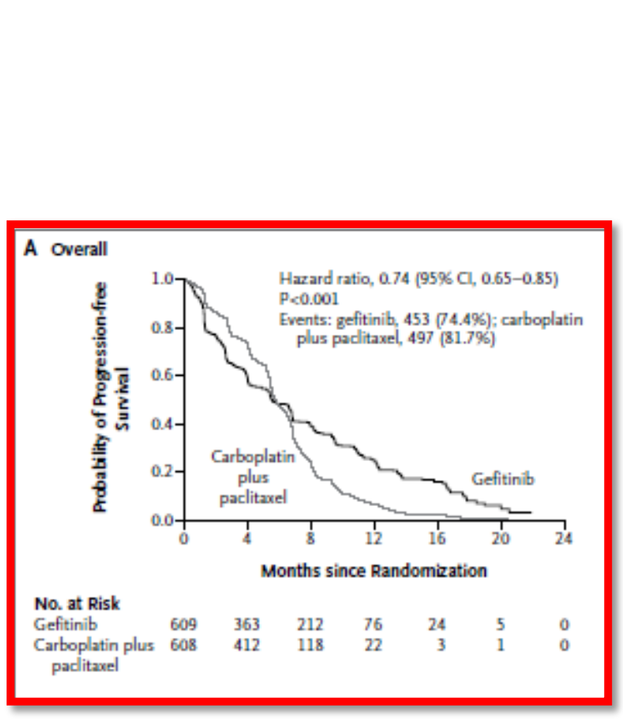
ESTABLISHED IN 1812 SEPTEMBER 3, 2009 VOL. 361 NO. 10

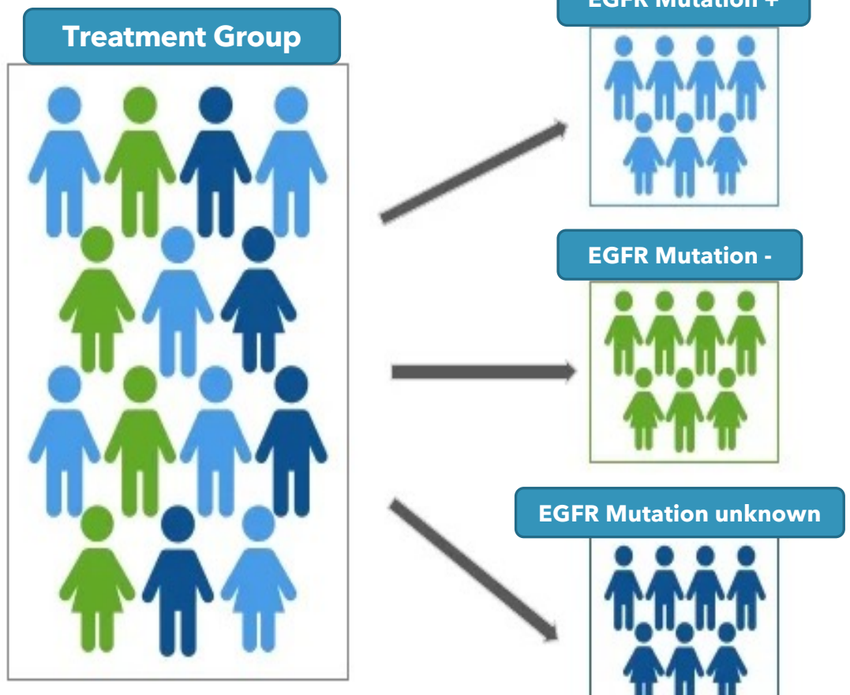
Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma

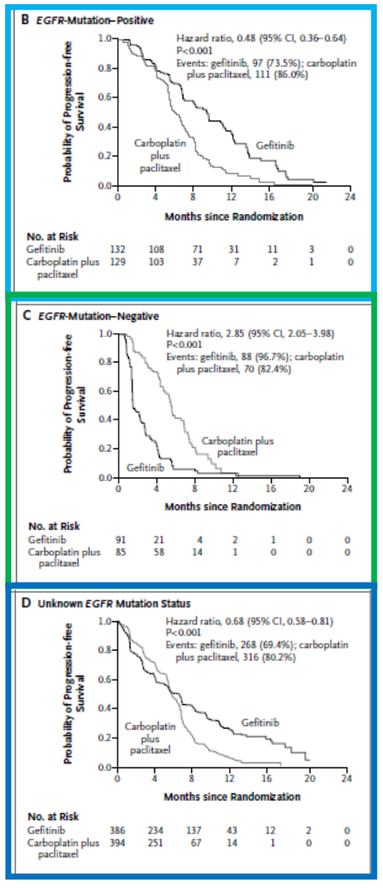
Tony S. Mok, M.D., Yi-Long Wu, M.D., F.A.C.S., Sumitra Thongprasert, M.D., Chih-Hsin Yang, M.D., Ph.D., Da-Tong Chu, M.D., Nagahiro Saijo, M.D., Ph.D., Patrapim Sunpaweravong, M.D., Baohui Han, M.D., Benjamin Margono, M.D., Ph.D., F.C.C.P., Yukito Ichinose, M.D., Yutaka Nishiwaki, M.D., Ph.D., Yuichiro Ohe, M.D., Ph.D., Jin-Ji Yang, M.D., Busyamas Chewaskulyong, M.D., Haiyi Jiang, M.D., Emma L. Duffield, M.Sc., Claire L. Watkins, M.Sc., Alison A. Armour, F.R.C.R., and Masahiro Fukuoka, M.D., Ph.D.

A real study

- Mok et al. Gefitinib or Carboplatin-Paclitasel in Pulmonary Adenocarcinoma. NEJM 2009; 361:947-957.







Non-proportionality Theorem

Non-proportionality Theorem

Theorem 1.

$$h(t) = \sum_{j=1}^{K-1} h_j - \frac{\sum_{j=1}^{K-1} h_{j^c}^{(K-1)} p_j S_C^*(t)^{h_j} + \left\{ \left(\sum_{j=1}^{(K-1)} h_j \right) - 1 \right\} p_K S_C^*(t)}{S_T^*(t)}$$

- $p_j = 100\% \Rightarrow h(t) = h_j$ *heterogeneous population*
- $h_k = 1$ *for all* $k's \Rightarrow h(t) = 1$ *ineffective treatment*

Non-proportionality Theorem

Theorem 3. The population hazard ratio function between treatment and control remains a constant only if the patient responses to treatment are homogeneous or the given treatment is ineffective to all treated subjects.

Our thought process..

Cause	Challenge	Solution
• Treating heterogeneous patients	• Differentiate various types of responders and non-responders	• Chance of response ≈ aggregated prevalence of each subgroup

PRIME+

PRIME+: P%-responder information embedded strategy:

- **Feature**: embed heterogeneous treatment response + delayed effect
  - Objective response, stable disease, progressive disease/non-response

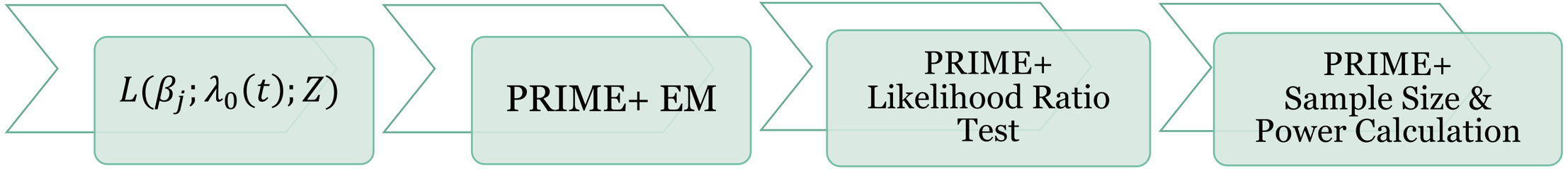
- **Aims**:
  - Study efficiency: Salvage power loss due to NPH patterns
  - Effect estimation: Detect subgroup-specific effect size

Model

- Mixture model:
  - heterogeneous treatment population
  - latent responder membership Z

$$\begin{cases} Z_i \mid i \in T \overset{i.i.d}{\sim} \textit{Multinomial}(p_1, p_2, \ldots, p_J) \\ Z_i \mid i \in C = 0 \end{cases}$$

PRIME+ Strategy



Re-design Nivolumab NSCLC Study By PRIME+

Re-design Nivolumab NSCLC Study

The Nivolumab NSCLC Study: Borghaei et al. NEJM 2015

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer

H. Borghaei, L. Paz-Ares, L. Horn, D.R. Spigel, M. Steins, N.E. Ready, L.Q. Chow, E.E. Vokes, E. Felip, E. Holgado, F. Barlesi, M. Kohlhäufl, O. Arrieta, M.A. Burgio, J. Fayette, H. Lena, E. Poddubskaya, D.E. Gerber, S.N. Gettinger, C.M. Rudin, N. Rizvi, L. Crinò, G.R. Blumenschein, Jr., S.J. Antonia, C. Dorange, C.T. Harbison, F. Graf Finckenstein, and J.R. Brahmer

ABSTRACT

BACKGROUND

Nivolumab, a fully human IgG4 programmed death 1 (PD-1) immune-checkpoint-inhibitor antibody, disrupts PD-1–mediated signaling and may restore antitumor immunity.

METHODS

In this randomized, open-label, international phase 3 study, we assigned patients with nonsquamous non–small-cell lung cancer (NSCLC) that had progressed during or after platinum-based doublet chemotherapy to receive nivolumab at a dose of 3 mg per kilogram of body weight every 2 weeks or docetaxel at a dose of 75 mg per square meter of body-surface area every 3 weeks. The primary end point was overall survival.

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Borghaei, Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111, or at hossein.borghaei@fccc.edu.

This article was published on September 27, 2015, at NEJM.org.

N Engl J Med 2015;373:1627-39.
DOI: 10.1056/NEJMoa1507643

Re-design Nivolumab NSCLC Study

Original Design: The Nivolumab NSCLC Study: Borghaei et al. NEJM 2015

- Nivolumab vs. Docetaxel in NSCLC
- Hybrid, simulation-based Design: 582 subjects to achieve 90% power

Re-design Nivolumab NSCLC Study

Original Design: The Nivolumab NSCLC Study: Borghaei et al. NEJM 2015

- Nivolumab vs. Docetaxel in NSCLC
- Hybrid, simulation-based Design: 582 subjects to achieve 90% power

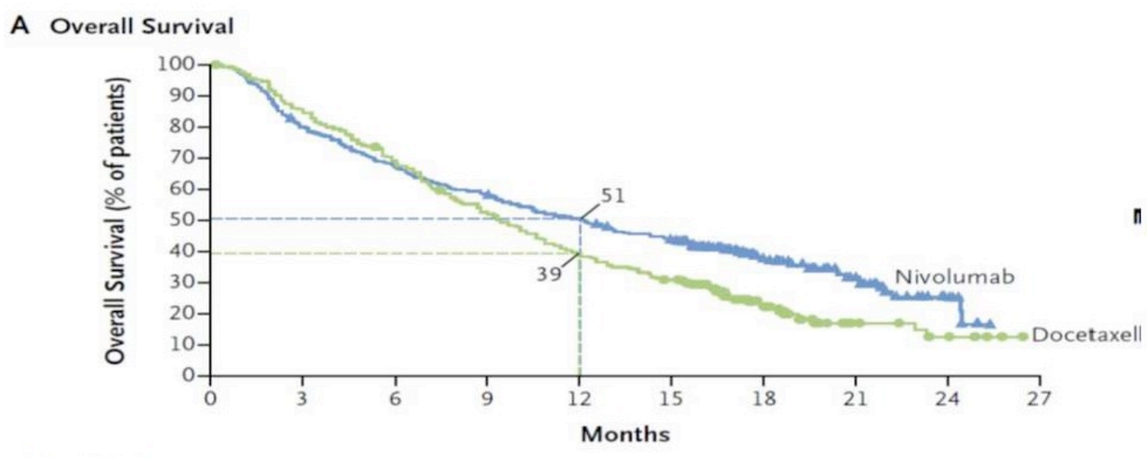
Re-design by PRIME+: 450 subjects to achieve 90% power

- $P_1 = 20\%, P_2 = 25\%, P_3 = 55\%$
- $\lambda_{OR} = 0.2$, $\lambda_{SD} = 0.52$
  - ORR = 20%; SDR = 25%; PR/NR = 55%
  - $\bar{\lambda}_T$=0.73 between Nivolumab vs Docetaxel
  - 20% OR + 25% SD + 55% NR $\Rightarrow \bar{\lambda}_T = 0.73$

Nivolumab Study Survival Patterns

Crossing hazards Pattern

Nivolumab versus Docetaxel in Advanced
Nonsquamous Non–Small-Cell Lung Cancer



No. at Risk										
Nivolumab	292	232	194	169	146	123	62	32	9	0
Docetaxel	290	244	194	150	111	88	34	10	5	0

Conclusions

Unique Features of our proposal:

APPLE, APPLE+: Delayed effect pattern

PRIME, PRIME+: Non-proportional hazards patterns

Cause  Solution

Advantages:

- Inference and treatment effect estimation:
  - Enhance efficiency
  - Provide clinical meaningful treatment effect estimation
  - Improve robustness

- Outline a strategy to mitigate occurrence of NPH patterns

Statistics in Medicine

Research Article

Received 29 March 2016, Accepted 2 October 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7157

Designing therapeutic cancer vaccine trials with delayed treatment effect

Zhenzhen Xu,[a*†] Boguang Zhen,[a] Yongsoek Park[b] and Bin Zhu[c]

Arming the immune system against cancer has emerged as a powerful tool in oncology during recent years. Instead of poisoning a tumor or destroying it with radiation, therapeutic cancer vaccine, a type of cancer immunotherapy, unleashes the immune system to combat cancer. This indirect mechanism-of-action of vaccines poses the possibility of a delayed onset of clinical effect, which results in a delayed separation of survival curves between the experimental and control groups in therapeutic cancer vaccine trials with time-to-event endpoints. This violates the proportional hazard assumption. As a result, the conventional study design based on the regular log-rank test ignoring the delayed effect would lead to a loss of power. In this paper, we propose two innovative approaches for sample size and power calculation using the piecewise weighted log-rank test to properly and efficiently incorporate the delayed effect into the study design. Both theoretical derivations and empirical studies demonstrate that the proposed methods, accounting for the delayed effect, can reduce sample size dramatically while achieving the target power relative to a standard practice. Copyright © 2016 John Wiley & Sons, Ltd.

RESEARCH ARTICLE

WILEY Statistics in Medicine

Designing cancer immunotherapy trials with random treatment time-lag effect

Zhenzhen Xu[1] | Yongsoek Park[2] | Boguang Zhen[1] | Bin Zhu[3]

[1]CBER, Food and Drug Administration, Silver Spring, Maryland

[2]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania

[3]DCEG, National Cancer Institute, Bethesda, Maryland

Correspondence

In some clinical settings such as the cancer immunotherapy trials, a treatment time-lag effect may be present and the lag duration possibly vary from subject to subject. An efficient study design and analysis procedure should not only take into account the time-lag effect but also consider the individual heterogeneity in the lag duration. In this paper, we present a Generalized Piecewise Weighted Logrank (GPW-Logrank) test, designed to account for the random

Package 'DelayedEffect.Design'

November 22, 2022

Title Sample Size and Power Calculations using the APPLE, SEPPLE, APPLE+ and SEPPLE+ Methods

Version 1.1.0

Date 2022-11-21

Author Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Boguang Zhen <Boguang.Zhen@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

Description Provides sample size and power calculations when the treatment time-lag effect is present and the lag duration is either homogeneous across the individual subject, or varies heterogeneously from individual to individual within a certain domain and following a specific pattern. The methods used are described in Xu, Z., Zhen, B., Park, Y., & Zhu, B. (2017) <doi:10.1002/sim.7157>.

Xu et al. Journal of Hematology & Oncology (2020) 13:20
https://doi.org/10.1186/s13045-020-0847-x

Journal of Hematology & Oncology

RESEARCH Open Access

Treating non-responders: pitfalls and implications for cancer immunotherapy trial design

Zhenzhen Xu[1*], Yongsoek Park[2], Ke Liu[1†] and Bin Zhu[3†]

Abstract

Background: Conventional trial design and analysis strategies fail to address the typical challenge of immune-oncology (IO) studies: only a limited percentage of treated patients respond to the experimental treatment. Treating non-responders, we hypothesize, would in part drive non-proportional hazards (NPH) patterns in Kaplan-Meier curves that violates the proportional hazards (PH) assumption required by conventional strategies. Ignoring such violation incurred from treating non-responders in the design and analysis strategy may result in underpowered or even falsely negative studies. Hence, designing innovative IO trials to address such pitfall becomes essential.

RESEARCH ARTICLE

Statistics in Medicine WILEY

Design for immuno-oncology clinical trials enrolling both responders and nonresponders

Zhenzhen Xu[1] | Bin Zhu[2] | Yongsoek Park[3]

[1]Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland

[2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

[3]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania

A typical challenge facing the design and analysis of immuno-oncology (IO) trials is the prevalence of nonproportional hazards (NPH) patterns manifested in Kaplan-Meier curves under time-to-event endpoints. The NPH patterns would violate the proportional hazards assumption, and yet conventional design and analysis strategies often ignore such a violation, resulting in underpowered or even falsely negative IO studies. In this article, we show, both empirically and analytically, that treating nonresponders in IO studies of inadequate size would

Package 'Immunotherapy.Design'

May 18, 2020

Title Study design for immunotherapy clinical trials

Version 1.1.0

Date 2020-05-18

Author Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

Description Perform sample size, power calculation and subsequent analysis for Immuno-oncology (IO) trials composed of responders and nonresponders.

Thank you

U.S. FOOD & DRUG ADMINISTRATION