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 Pittsburgh
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 Administration

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

Diminishing effect

Crossing hazards

Combination
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 HEMAT (Revlimid).

t* = 18
Motivating example

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

The Kaplan-Meier curves for progression-free survival: Study FIRMST (Revlimid).

Figure 1.
Motivating example
Motivating example
Theorem 1. Under fixed delayed scenario, the optimal weights $W_j^* = \arg \max \{ \text{Pow}(w_j) \}$ need to satisfy that $W_j^* \propto \log \{ \lambda(t_j) \}$.

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

\[
W^*(t) = \begin{cases} 0, & t < t^* \\ 1, & t \geq t^* \end{cases}
\]
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:

  \[ \textit{FDA Chief Scientist Publication Award:} \text{An exceptional manuscript with immediate impact that may speed availability of cancer therapies} \]

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

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

Figure 1. Kaplan–Meier curves for progression-free survival: Study 001 (Revlimid).
Motivating example

**The NEW ENGLAND JOURNAL of MEDICINE**

**Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma**

![Graph showing Kaplan-Meier curves for progression-free survival.](image)

**Figure 1.** Kaplan-Meier curves for progression-free survival: Study EP1ST (Revlimid).

<table>
<thead>
<tr>
<th>Time</th>
<th>Treatment</th>
<th>MPT</th>
<th>NPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>547 (0)</td>
<td>535 (0)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>304 (100)</td>
<td>319 (104)</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>170 (127)</td>
<td>218 (84)</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>78 (106)</td>
<td>135 (104)</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>5 (229)</td>
<td>19 (239)</td>
</tr>
<tr>
<td>72</td>
<td></td>
<td>0 (237)</td>
<td>0 (215)</td>
</tr>
</tbody>
</table>

\[ T_1 = 18 \quad T_2 = 33 \]
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
Theorem 2. Under random delayed scenario, the optimal weights $W_j^* = \text{argmax}\{\text{Pow}(w_j)\}$ need to satisfy that $W_j^* \propto F_*(t_j)$.

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

$$W^*(t) = F_*(t)$$
Generalized Piecewise Weighted Logrank Test

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

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

The NEW ENGLAND JOURNAL of MEDICINE

Lenalidomide and Dexamethasone in Transplant-Ineligible Patients with Myeloma

\[ W_{(t \leq T_1)} = 0 \]
\[ W_{(T_1 < t \leq T_2)} = \frac{(t - T_1)}{(T_2 - T_1)} \]
\[ W_{(t > T_2)} = 1 \]

Figure 1. Kaplan–Meier curves for progression-free survival: Study E1607 (Revlimid).
Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True parameter setting</strong>: Fixed scenario with $t^* = 6$</td>
<td></td>
</tr>
<tr>
<td>PW-Logrank $t^* = 6$</td>
<td>79%</td>
</tr>
</tbody>
</table>
## Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>True parameter setting: Fixed scenario with ( t^* = 6 )</td>
<td></td>
</tr>
<tr>
<td>PW-Logrank ( t^* = 6 )</td>
<td>79%</td>
</tr>
<tr>
<td>PW-Logrank ( t^m = 1 )</td>
<td>63%</td>
</tr>
<tr>
<td>PW-Logrank ( t^m = 11 )</td>
<td>64%</td>
</tr>
</tbody>
</table>
## Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True parameter setting:</strong> Fixed scenario with ( t^* = 6 )</td>
<td></td>
</tr>
<tr>
<td>PW-Logrank ( t^* = 6 )</td>
<td>79%</td>
</tr>
<tr>
<td>PW-Logrank ( t^m = 1 )</td>
<td>63%</td>
</tr>
<tr>
<td>PW-Logrank ( t^m = 11 )</td>
<td>64%</td>
</tr>
<tr>
<td>GPW-Logrank ([T_1, T_2] = [1,11])</td>
<td>76%</td>
</tr>
<tr>
<td>GPW-Logrank ([T_1, T_2] = [1,9])</td>
<td>76%</td>
</tr>
<tr>
<td>GPW-Logrank ([T_1, T_2] = [3,9])</td>
<td>78%</td>
</tr>
</tbody>
</table>
Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True parameter setting:</strong> Random scenario with $[T_1, T_2]=[3,9]$</td>
<td></td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^<em>, T_2^</em>]=[3,9]$</td>
<td>80%</td>
</tr>
</tbody>
</table>
### Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True parameter setting</strong>: Random scenario with $[T_1, T_2]=[3,9]$</td>
<td></td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^<em>, T_2^</em>]=[3,9]$</td>
<td>80%</td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^m, T_2^m]=[1,9]$</td>
<td>79%</td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^m, T_2^m]=[3,11]$</td>
<td>79%</td>
</tr>
</tbody>
</table>
## Advantage of GPW Logrank test vs PW Logrank test

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True parameter setting:</strong> Random scenario with $[T_1, T_2]=[3,9]$</td>
<td></td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^<em>, T_2^</em>]=[3,9]$</td>
<td>80%</td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^m, T_2^m]=[1,9]$</td>
<td>79%</td>
</tr>
<tr>
<td>GPW-Logrank $[T_1^m, T_2^m]=[3,11]$</td>
<td>79%</td>
</tr>
<tr>
<td>PW-Logrank $t^m = 1$</td>
<td>66%</td>
</tr>
<tr>
<td>PW-Logrank $t^m = 11$</td>
<td>69%</td>
</tr>
</tbody>
</table>
APPLE+, SEPPLE+

Generalized Piecewise Weighted Logrank Test

• APPLE $\xrightarrow{\text{}}$ APPLE+
• SEPPLE $\xrightarrow{\text{}}$ SEPPLE+
How to deal with general NPH Patterns?

- **Delayed effect**
- **Diminishing effect**
- **Crossing hazards**

**Combination**

*Nivolumab versus Docetaxel in Advanced Non-Squamous Non-Small-Cell Lung Cancer*

*Nivolumab plus ipilimumab in advanced melanoma without BRAF Mutations*

*Sipuleucel-T Immunotherapy for 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 $\leftrightarrow$ NPH?
A real study

- A limited percentage of treated subjects respond whereas others don’t
- Are we treating heterogeneous patients \(\iff\) NPH?
A real study

Non-proportionality Theorem
Theorem 1.

\[ h(t) = \sum_{j=1}^{K-1} h_j - \frac{\sum_{j=1}^{K-1} h_j^{(K-1)} p_j S^*_C(t)^{h_j} + \left\{ \sum_{j=1}^{K-1} h_j \right\} p_K S^*_C(t)}{S^*_T(t)} \]

- \( p_j = 100\% \Rightarrow h(t) = h_j \) heterogeneous population
- \( h_k = 1 \text{ for all } k' \text{s} \Rightarrow h(t) = 1 \) 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...

- Treating heterogeneous patients
- Differentiate various types of responders and non-responders
- Chance of response $\approx$ aggregated prevalence of each subgroup
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 $Z$

$$
\begin{align*}
Z_i \mid i \in T & \overset{i.i.d}{\sim} \text{Multinomial}(p_1, p_2, ..., p_J) \\
Z_i \mid i \in C & = 0
\end{align*}
$$
PRIME+ Strategy

\[ L(\beta_j; \lambda_0(t); Z) \]

PRIME+ EM

PRIME+ Likelihood Ratio Test

PRIME+ Sample Size & Power Calculation
Re-design Nivolumab NSCLC Study By PRIME+
Re-design Nivolumab NSCLC Study

The Nivolumab NSCLC Study: Borghaei et al. NEJM 2015
**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

![Graph showing survival patterns for Nivolumab and Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer](image)

- **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

- **Overall Survival**
  - Y-axis: Overall Survival (% of patients)
  - X-axis: Months

- Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer
Conclusions
Unique Features of our proposal:

APPLE, APPLE+: Delayed effect pattern
PRIME, PRIME+: Non-proportional hazards patterns
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
Research Article

Designing cancer immunotherapy trials with random treatment time-lag effect

Zhenzhen Xu, Boguazhen, Yongsook Park, and Bin Zhu

Abstract

An immune system against cancers has emerged as a powerful tool in oncology during recent years. Instead of promoting a tumor or destroying it with radiation, therapeutic cancer vaccines, a type of cancer immunotherapy, induces 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. This delay can lead to underpowered tests and Type II error. In this paper, we propose two innovative approaches for sample size and power calculation using the placebo-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 effects, can provide more accurate sample size estimations, achieving the target power relative to a standard practice. Copyright © 2016 John Wiley & Sons, Ltd.

RESEARCH ARTICLE

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

Zhenzhen Xu, Yongsook Park, and Bin Zhu

A typical challenge facing the design and analysis of immuno-oncology (IO) trials is the presence of nonproportional hazards (NPH) patterns manifest in Kaplan-Meier curves that violates the proportional hazards (PH) assumption required by conventional strategies. Ignoring such violation, accrued from treating nonresponders in the design and analysis stage may result in undervalued or even falsely negative results. Hence, designing innovative IO trial to address such pitfalls becomes essential.
Thank you