

**Public Workshop: Oncology Clinical Trials
in the Presence of Non-Proportional Hazards**

The National Press Club • Washington, DC
February 5, 2018

Discussion Guide

The pace of development and approvals for novel cancer treatments has been steadily increasing, representing a shift in potential treatment options that could transform patient care. In the last decade alone, the pipeline of oncology drugs in clinical development has expanded by 45%¹ with 68 novel cancer therapies launched globally between 2011 and 2016. In 2017, the U.S. Food and Drug Administration (FDA) approved 13 novel cancer therapies, as well as 7 new uses for previously approved cancer treatments.² At the same time, the types of cancer therapies developed and approved are changing, representative of new mechanisms of action or entirely new classes of drugs. Immunotherapies, for example, use novel agents to target cancer, including immune checkpoint inhibitors, monoclonal antibodies, and vaccines, among others. These and other new types of cancer therapies are an increasing share of the drug development pipeline,³ with first-in-class products representing approximately 85% of the current cancer pipeline.⁴

While many of these new treatments show great promise, there are ongoing challenges with oncology clinical trial designs and analysis of treatment effects for cancer therapies that may impact our understanding of investigational drugs.

Traditionally, randomized clinical trials for cancer products have considered time-to-event endpoints such as progression-free survival and overall survival as the primary outcome measure in trial design. The most commonly used statistical tools to analyze these time-to-event endpoints are the Kaplan Meier survival plot, the log-rank test, and the Cox proportional hazard regression model. The performance of these tests largely depends on the proportional hazards (PH) assumption – that the hazard ratio is constant over time. A hazard ratio is a comparison of the probability of events in a treatment group to the probability of events in a control group. However, the proportionality assumption generally does not hold true in cases where there is delayed treatment effect, diminishing treatment effect, long term survival, treatment effect only in a subgroup, or when the survival curves cross each other. In such cases, summarizing the treatment effect based on the hazard ratio may not yield the most accurate analysis, and the clinical trial designs based on this assumption are likely to be under-powered to detect treatment differences.

These distinctive characteristics in the survival curves are observed in many clinical trials across various cancers and drug products, and are especially prevalent for immunotherapies. To better capture and characterize these properties, potential new endpoints, modifications to traditional endpoints, new statistical methods, and new clinical trial design paradigms are all needed. Recognizing a collaborative need to address these issues, FDA formed a working group with pharmaceutical companies to: 1) systematically review assumptions under different statistical methods used in time-to-event analyses; 2) identify appropriate statistical tests under different non-proportionality conditions; and 3) identify summary measures for describing treatment effect in randomized clinical trials.

This workshop, convened by the Duke-Margolis Center for Health Policy under a cooperative agreement with FDA, will serve as a public forum for presenting and discussing the cross-pharma working group's progress-to-date. Discussion will encompass the development and application of innovative oncology clinical trial designs, alternative statistical methods for evaluating treatment effects of time-to-event endpoints, and potential outcome measures that more accurately capture treatment effect.

Session I: Perspectives on Potential Limitations of Statistical Plans Based on Assumption of Proportional Hazards Over Time

This session will provide an overview of the limitations and challenges that accompany most commonly used clinical trial designs and analytical plans. These include challenges and limitations with current statistical methods for summarizing measures to describe treatment effect, and for analyzing relative treatment effect when the difference between experimental treatment and a control (standard of care) treatment is not constant over time. Presentations in this session will highlight recent examples of these challenges and how they led to the formation of the cross-pharma working group.

Session II: Analysis Methods and Simulations: Addressing Non-Proportional Hazards

In this session, discussion will focus on alternative statistical methods for analyzing time-to-event data in the presence of non-proportional hazards (NPH). The performance of commonly used statistical tools for analyzing time-to-event endpoints heavily depends on the PH assumption. When the PH assumption is violated, these tools have limited interpretation for treatment effect. Therefore, new statistical tools are required to better detect and quantify the treatment benefit under conditions of NPH.

In identifying new methods for analyzing time-to-event data, the cross-pharma working group first examined widely used methods for hypothesis testing and estimation in the presence of NPH, such as log-rank based tests, restricted mean survival time, and the piecewise Cox model. They determined that a new method was needed to analyze these time-to-event data and have proposed a new combination test based on Fleming-Harrington weighted log-rank statistics. The proposed methodology provides robust inference for a large class of NPH scenarios of interest.

In order to demonstrate the validity of this new combination test, the working group has performed a simulation exercise to compare currently used methods with the groups' new test by analyzing operating characteristics (e.g., type I error, power). This simulation exercise includes a wide range of scenarios where this new method might apply, including in the presence of different types of NPH (e.g. delayed effect, crossing hazard, and diminishing treatment effect), different degrees of censoring observations, and different alternatives for treatment effect (e.g. eNull, PH, and different types of NPH). Based on the overall simulation results, the new combination test performs reasonably well under many different scenarios and clearly has greater power to test treatment difference than the other methods under delayed treatment effect and crossing hazard scenarios.

Discussion Questions:

- Are the limitations and advantages of each of the methods presented easily understood? If not, what information would help clarify any outstanding questions?
- Given the simulations performed, are the assumptions for the simulations appropriate?
- Should reporting of clinical benefit be dependent on the presence of non-proportional hazards (i.e., tailored to the specific data set?)

- In the presence of non-proportional hazards, should we consider using more than one summary measure to capture treatment benefit?
- Is the proposed combination test a robust and appropriate test to use when there is uncertainty of non-proportional hazards? Are there drawbacks to the proposed method?
- By design, the combination test will potentially down-weight early or late events. Is this an acceptable approach?
- Are there any other alternative methods that should be considered?

Session III: Retrospective Application of Novel Analysis Methods in Completed Trials

In this session, the working group will highlight how the proposed combination test can be applied in real-world settings and how the test's results compare to those from statistical tools currently in use. Utilizing case studies from four completed cancer clinical trials, presenters will illustrate why this new approach may be more appropriate for analyzing time-to event data in the presence of NPH. The trials were selected based on visual inspection of the Kaplan Meier curves, with the goal of representing different patterns of NPH.

The following studies are included:

- The INO-VATE study comparing inotuzumab versus standard chemotherapy in patients treated with relapsed or refractory acute lymphoblastic leukemia
 - NPH pattern: delayed effect and long-term remission
- Phase 3 study comparing Ipilimumab 10 mg/kg vs ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma
 - NPH pattern: delayed effect and long-term remission
- A two-arm randomized phase 2 study of mitoxantrone and prednisone (MP) plus cixutumumab or ramucirumab in patients with metastatic castration-resistant prostate cancer (mCRPC)
 - NPH pattern: diminishing treatment effect
- The IPASS study comparing gefitinib versus carboplatin plus paclitaxel in patients with advanced pulmonary adenocarcinoma
 - NPH pattern: crossing survival curves

In all four studies, progression free survival or overall survival was a primary or key secondary endpoint. The log-rank test, the Fleming-Harrington class of weighted log-rank tests, the combination test (using Fleming-Harrington weights), and restricted mean survival time were used to compare the two treatment arms. The results from these comparisons were generally consistent with the findings from the simulation study and showed robustness of the combination test under different types of NPH scenarios.

Discussion Questions:

- For each of the case examples presented, was the optimal method used to evaluate relative treatment effect? If not, what would have been a better approach?
- Based on the test cases discussed, does it make sense from both a statistical and clinical perspective to use the combination test as an alternative to the log-rank test in certain clinical settings? Are there specific clinical settings where it would not be appropriate to use the proposed combination test?

- Are there any concerns with the assumptions made in applying the proposed combination test in each of these cases?
- Are there any elements of statistical analysis that were not adequately considered?

Session IV: Considerations for Improving Future Trial Designs

In this session, discussion will center on how to set up and design future studies employing the newly proposed combination test. An important consideration when designing these studies is that researchers may not know in advance if the study will involve proportional or non-proportional hazards. By utilizing different case study examples, the working group will outline a number of strategies for designing a study in the context of this uncertainty. This will include recommendations with regard to power, sample size, and follow-up time that can help guide researchers as they encounter different statistical scenarios.

Discussion Questions:

- Would sponsors consider using the proposed combination test when designing a future clinical trial? Why or why not?
- If there is no prior information that signals non-proportional hazards, would sponsors still consider using the proposed combination test?
- Given the new method proposed, how should the results be summarized? Are there other options to consider for summarizing treatment effect?
- How can the FDA develop new guidelines for using a combination test for the primary analysis?
- How important is long-term follow-up for the primary analysis as well as for subsequent supportive analyses and label updates?
- How should we pre-specify description of benefit? Potential options include but are not limited to:
 - Cox and weighted Cox estimates
 - Piecewise proportional hazards over time
 - Milestone estimate, including follow-up after primary analysis
 - Restricted mean survival time to express measure of life-expectancy gain
- When conducting a clinical trial, should there be early evaluations of efficacy or safety (including overall survival)?
- When is it acceptable to ignore an early unfavorable trend? (e.g. timing, degree). How would you describe this process in your statistical package when submitting to FDA?
- Would you consider an alternative method after reviewing interim analysis results? Are there drawbacks to using interim data?

Session V: Implications for the Broader Stakeholder Community

Given the potential broader application of alternative statistical analyses for assessing NPH, stakeholders may need to further explore the impact these new approaches may have on other stakeholders' decision making, as well as barriers to implementation or uptake of the new approaches. This session will consider these issues, including the need for training both researchers and reviewers in these techniques. Panelists will highlight potential impacts on patient care, including challenges for physicians in interpreting results for clinical decision-making, and whether payers will alter reimbursement decisions based on these results. Discussion will also highlight any outstanding questions that must be addressed following the workshop. These may include issues relating to the

proposed test itself, its regulatory acceptability among international regulatory agencies, challenges with implementation in clinical trials, and how this implementation may impact all stakeholders. Finally, the panel will outline potential next steps that must take place to move this initiative forward.

Discussion Questions:

- What are the important issues that regulators, clinicians, and other stakeholders should consider as new alternative methods of analysis are proposed?
- Are the methods presented for analyzing time-to-event data clinically interpretable? What are the potential challenges?
- How might the intended use of information related to treatment effect (e.g., for supporting clinician and patient decision-making or payer coverage decisions) be affected by changes to the traditional design or analytical approaches?
- How do clinicians understand and explain to their patients the implications of these time-to-event data for treatment benefit? How might the proposed methods affect that process?
- How might patients use or interpret the results given uptake of the proposed methods?
- How might information and clinical evidence resulting from the use of the proposed methods change how payers make formulary or coverage decisions? What additional information about the combined test or how to interpret its results might be helpful?
- Will these alternative methods be acceptable to international regulators? What regulatory harmonization or uptake challenges could arise?
- How should the results of these trials be presented in product labels?

¹ https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/global-oncology-trends-2017.pdf?_id=1516825034305

²

<https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ReportsBudgets/UCM591976.pdf>

³ <https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/immunotherapy/what-is-immunotherapy.html>

⁴ http://www.analysisgroup.com/uploadedfiles/content/insights/publishing/the_biopharmaceutical_pipeline_report_2017.pdf