Statistics to Aid Decision Making for Early-Phase Cancer Trials with Time-to-Event Endpoints

Shaoyi Li
Celgene Corporation, Berkeley Heights, NJ 07922

Key words: randomized reference, projected Kaplan-Meier curve, simulated reference curve, early-phase cancer trials, time-to-event endpoints, single-arm

ABSTRACT

As shown in many publications, the high failure rate of confirmatory trials may be due to the discordance in the endpoints adopted in Phase I/II trials. The commonly-used endpoint in early-phase cancer trials is the tumor response as opposed to survival in late-phase. While a robust tumor response is expected to translate into survival benefit, it may not always be true. The present study was designed to address this issue by adding a survival endpoint to the dose expansion phase. With limitation of small budget, a single-arm design is preferred. In this case, two issues need to be addressed: (a) regulatory guidance indicates that the interpretation of time-to-event single-arm data without a randomized reference is problematic; (b) the trial is too short which leads to a lot of survival data being censored when it comes to the decision point. Based on discussions with the clinical team, a simulation approach was proposed to project the Kaplan-Meier curve longer term. A reference curve was also simulated based on the input from key opinion leaders and publications. The final simulation result is helpful to aid decision making for early-phase cancer trials with time-to-event endpoints such as survival.

1. INTRODUCTION

The high failure rate of confirmatory trials in oncology may be due to the discordance in the endpoints for the different development phase of an oncology drug. Usually, the primary efficacy endpoint is the objective response rate (ORR) in early-phase and the overall survival (OS) in late-phase. Numerous clinical trials' results revealed that a robust tumor response may not always translate to a survival benefit unfortunately.

The other aspect pertaining to the topic under discussion is that, for early-phase cancer trials especially, with limitation of small budget, a single-arm design is preferred. In the FDA's Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (2007), the FDA has sometimes supported ORR and response duration observed in single-arm studies as substantial evidence supporting accelerated approval in settings where there is no available therapy and where major tumor regressions can be presumed to be attributed to the tested drug. However, it is clearly stated in the FDA Guidance that single-arm trials do not adequately characterize time-to-event (TTE) endpoints such as survival, time to progression (TTP), or progression free survival (PFS). The EMA's Oncology Guidance also has the similar position that the interpretation of TTP/PFS data without a randomized reference is problematic. There is no disease control rate (DCR) mentioned. As a matter of facts, DCR may not be a good short-term surrogate endpoint for survival (e.g., EVOLVE-1 HCC Phase III trial (Zhu 2014). The OS p-value=0.675 while DCR p-value=0.01).

Section 2 will show the dilemma of endpoint choices in early-phase oncology in general through some typical examples which support the position of the Health Agencies on single-arm TTE endpoints. Section 3 will share one example of statistics using simulation for a reference arm based on the median OS estimate from the key opinion leaders (KOL) on the tumor type of interest. The simulation results are to aid decision making when running into the situation of single-arm survival. Section 4 will discuss some potential future research for the topic.

2. THE DILEMMA

The issue of a robust tumor response may not always translate to a survival benefit has been discussed in the project team. The survival endpoint has been built in the dose expansion phase of the present study based on the maximum tolerated dose (MTD) identified from the dose escalation phase. A single-arm design is preferred with the limitation of a small budget for an early-phase trial. The following are the results (masked), where XXX is the study drug and YYY and ZZZ are competitors.

	XXX* (30 mg)	**XXX* (15 mg)**	**YYY* (2nd line)**	**ZZZ* (2nd line)**
N (All)	25[1]	28[1]	265[2]	365[2]
ORR (%) [95% CI]	8 [1, 25]	3.5 [0.1,18.1]	12 [6.5,14]	2.2
DCR (%) [95% CI]	40 [24,65]	49 [30, 70]	60 [55,66]	55[51,61] p=0.014
Median TTP (mo)	3.7	5.2	4.2	3.0
Median PFS (mo)	3.5	3.6	N/A	N/A
Median OS (mo)	5.8	6.9	9.5	7.5

[1]All Treated; [2]Intent-to-Treat; *Data has been masked for confidentiality. ORR: CR+PR. DCR: CR+PR+SD.

Caution must be taken for the above side-by-side comparisons among different studies especially from different companies for single-arm setting. The following are good examples.

Example 1: RADIANT-2 (Pavel, Lancet 2011)

The results of median PFS, same trial data but difference reviewers, show that 16.4 (months) (treatment) vs. 11.3 (placebo) from central reviewer while 12.0 (treatment) vs. 8.6 (placebo) from investigator. Therefore, 16.4/12=1.367 –> about 37% inflation for treatment and 11.3/8.6=1.31 –> 31% inflation for placebo. Simply, different reviewers may have very much different PFS estimates based on the same radiological data. Naturally, the results from different reviewers on different compounds from different companies are very likely not going to be comparable in a single-arm setting.

Example 2: Sorafenib HCC 1L Phase III trials (Llovet, 2008; Cheng, 2009)

The ratios of median OS with different regions are 10.7/6.5=1.65 for sorafenib which is 65% inflated while 5.5/2.8=1.96 for placebo which is 96% inflation. It seems that the different regions of the world do have impact on the survival benefit using sorafenib and placebo as single-arms respectively. The TTP also demonstrates the similar results. However, the hazard ratios (HR) are very much the same with almost 45% improvement

for survival relative to the placebo for two independent trials which led to the approval of the NDA.

Study locations	NA, SA, Europe, Australia			Asia-Pacific (China, Taiwan, South Korea)		
	Sorafenib	Placebo	HR	Sorafenib	Placebo	HR
OS (months)	10.7	7.9	0.69	6.5	4.2	0.68
TTP (months)	5.5	2.8	0.58	2.8	1.4	0.57

It reveals an important observation: if these two trials were without the placebo as a control, it probably wouldn't have made for a positive conclusion. Without a reference/ control arm, it is difficult to tell how good is good and how bad is bad! The following two additional examples would provide two pieces of factual evidence.

Example 3: Sorafenib HCC 1L Phase III vs. Phase II (Llovet, 2008; Abou-Alfa, 2006)

This is a successful case moving from single-arm Phase II to Phase III trial. The results show that with 10.7/9.2=1.16 for sorafenib median OS which is 16% inflation while 5.5/5.5 stays the same for sorafenib TTP.

Study	**NA, SA, Europe, Australia; Phase III**			**Phase II**
	Sorafenib	Placebo	HR	Sorafenib
OS (months)	10.7	7.9	0.69	9.2
TTP (months)	5.5	2.8	0.58	5.5

Does it mean that sorafenib got lucky? YES, in terms of using single-arm Phase II OS result for the decision of moving to Phase III. And NO, because two Phase III trials independently confirmed the OS statistical significance (see example 2).

Example 4: Everolimus HCC 2nd-Line Phase III vs. Phase II (Zhu, 2014; Pavel, 2011)

This is a failed case moving from single-arm Phase II to Phase III trial. The median OS results show that 7.6/8.4=0.90 for everolimus which is only 10% deflation.

	Everolimus Phase III			**Phase II**
	Everolimus	Placebo	HR	Everolimus
OS (months)	7.6	7.3	1.05	8.4

Does it mean that everolimus got unlucky? YES, in terms of using single-arm Phase II OS result for the decision making of moving to Phase III. And NO, because the Phase III HR wasn't what was expected. The Phase II trial didn't provide a good hint in that regard. The decision making based on such a single-agent Phase II trial for Phase III is quite risky. Many other pivotal trials also failed in the similar way!

The above examples strongly support the position of the Health Agencies on single-arm TTE endpoints. It reveals that a single-arm design with survival endpoint(s) will not help us much in providing the hint for the survival benefit for advancing early-phase compounds to late-phase. As a result, we run into such a dilemma that a randomized reference arm is warranted for even a Phase II trial if TTE endpoints are important for the compound of interest. On the other hand, the budget would be literally doubled and the trial would be longer.

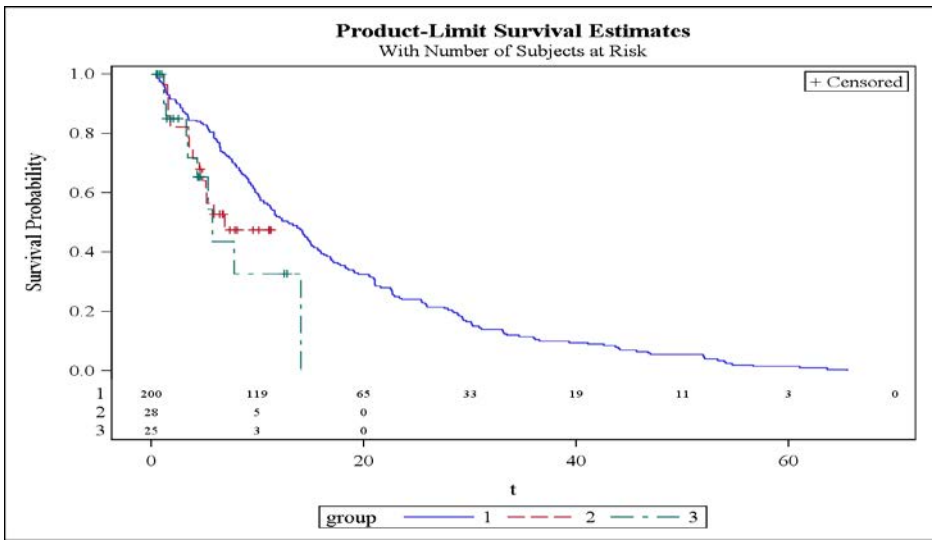
3. STATISTICS TO AID DECISION MAKING

How we overcome the dilemma that a randomized reference arm built within the trial would double the budget, but single-arm trials do not adequately characterize TTE endpoints. A historical control in an aggregate level may be difficult to justify either with the similar reasons as shown in the previous examples. Besides, the early-phase trial is too short which leads to a lot of survival data being censored when it comes to the decision point. We proposed the following:

- A reference simulated curve based on the input from the key opinion leaders (KOL) and publications.
- A simulation approach to project the K-M curve longer term for two scenarios:
    a. Survival Curves (real data) vs. Reference Curve (simulated)
    b. Survival Curves (semi-simulated) vs. Reference Curve (simulated)

a. Survival Curves (real data) vs. Reference Curve (simulated)

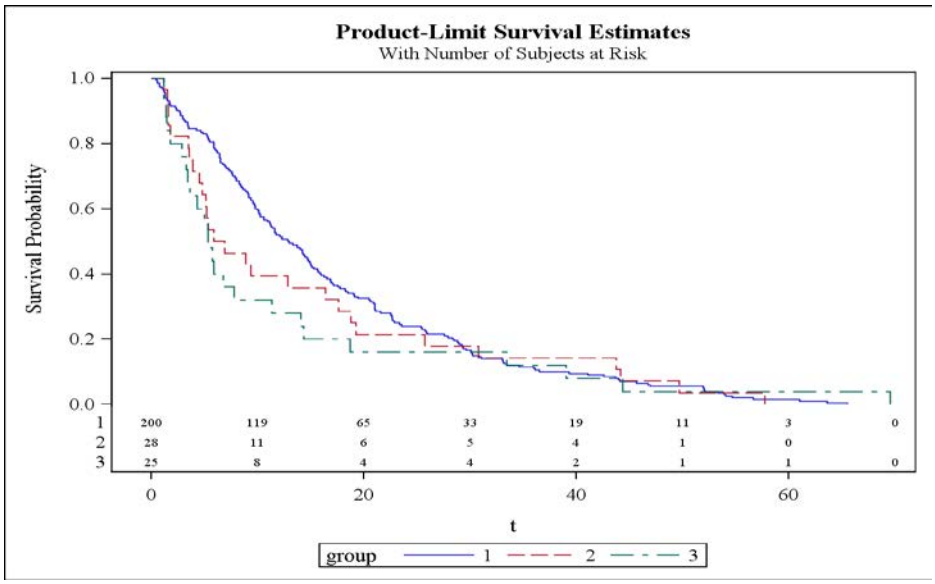
Group 1: Reference (median OS=12 (month)); Group 2: 15mg; Group 3: 30mg



Reference survival curve is simulated with n=200 subjects by using best scenario of 12 months median survival based on the input from the key opinion leaders (KOL) and publications. Survival curves for 15 mg and 30 mg arms are created median OS (95%CI) (in month): Group 2=6.9 (4.5, NA) and Group 3=5.8 (3.5, 14.1), respectively.

b. Survival Curves (semi-simulated) vs. Reference Curve (simulated)

Group 1: Reference (median OS=12 (month)); Group 2: 15mg; Group 3: 30mg.



Reference survival curve is simulated the same way as for the previous case. Because we have many censored survival data at the decision point, the simulation approach was performed based on the best scenario from KOL just for those censored patients. An exponential distribution was assumed. The median OS (95%CI) (mo): Group 2=6.4 (4.5, 16.4); Group 3=5.4 (3.5, 11.3).

For both cases, the graphs show that very unlikely the treatment is going to make it for the best scenario even though the 95% confidence interval of the median OS overlapped with that of the reference. The final simulation result is helpful to aid decision making for early-phase cancer trials with a survival endpoint such as OS, and can be extended to other TTE endpoints such as PFS, or TTP.

4. DISCUSSIONS

For the simulated reference, a lot of existing published information could also be processed and built into the simulation. Besides, we could be even more creative by fully utilizing the idea of big data of the standard care/control collected from different companies over years. They are subject-level data but with de-identification. We probably could “filter” and “normalize” the data so that they can be used as a reference arm at least sensible enough for early-phase single-arm trials. In practice, if we have to use single-arm study for the TTE endpoint, OS is a better choice over PFS and TTP, if possible, because OS is less subject to TTE censoring and tumor evaluations, still sensitive to region though. More research are warranted in this area.

Acknowledgements

Many thanks to Dr. Kristen Hege for the insightful discussions, to Dr. Gary Cline for his guidance, and to Dr. Daniel Meng, Statistical Contractor, for his hard working on the simulation programming.

References

Abou-Alfa, G. K.,, et. Al. (2006). Phase II Study of Sorafenib in Patients With Advanced Hepatocellular Carcinoma. *J Clin Oncol*, 24:4293-4300.

Cheng, A. L., et. al. (2009). Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol*, 10: 25–34.

Llovet, J. M., et. al. (2008). Sorafenib in Advanced Hepatocellular Carcinoma. The New *England Journal of Medicine*, 359:378-90.

Pavel, M. E., et. al. (2011). Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, Phase III study. *Lancet Oncol*, 378(9808), 2005-12.

Zhu, A. X., et. al. (2014). Effect of Everolimus on Survival in Advanced Hepatocellular Carcinoma After Failure of Sorafenib The EVOLVE-1 Randomized Clinical Trial, *JAMA*, 312(1):57-67.

Zhu, A. X., et. al. (2011). Phase ½ Study of Everolimus in Advanced Hypatocellular Carcinoma. *Cancer*, Volume 117 (22): 5094-5102

Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (May 2007). U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER).