How to Make a Better-Informed Go-No-Go Decision in Oncology Trials Using Tumor Measurement Data and Beyond

Bob Zhong[1], Bo Gao, Jing-ou Liu, Debra Mclntyre, CG Wang

Abstract

In this paper, we discussed new methods to summarize clinical efficacy data. The method is called "**T**rajectory and their corresponding non-missing **P**roportions **P**lot (TPP). The method focuses on the conditional mean response, condition on that the data are non-missing, and the proportion of non-missing, overtime. This graph is proposed for all trials to assess the treatment effect and dropout pattern across all time points. These two parameters are jointly used to access the treatment effect to answer a particular question about effectiveness that reduces the ambiguity of phrasing in federal regulations. It will be seen that it requires both parameters to conclude that one treatment is better than other at a particular time point. The first parameter is obvious, while the second parameter is from the belief that the proportion of non-missing from the more effective treatment (*A)* should not be less than that of the less effective treatment (*B)* to conclude that *A* is better than *B*. Another way to put this is: to conclude *A* is better than *B*, we should not know less about *A* than *B*.

From the parameters defined above, the missing proportions are an integral and intrinsic part of our data and are used in assessing non-missing probability. The non-missing proportion overtime reflects the reality in practice and is a very important information for stakeholders.

The paper focuses on using the approach in making go-no-go decision in oncology proof-of-concept trials.

Key Words: Tumor measurement data, go-no-go decision, RECIST.

---

1. contact information: chongxin.zhong@regeneron.com

1. Introduction

In the compound development process in the oncology therapeutic area, after a compound has passed the pre-clinical phase and dose-finding phase, the common approach is to conduct one or more proof-of-concept trials. The purpose of these trials is to assess if the compound shows early efficacy signals in some diseases hence indications can be pursued in these disease areas. Once these disease areas are identified, phase 3 trials are commonly launched. The typical difficulty during or at the end of the proof-of-concept trials is the uncertainty of deciding go-no-go for phase 3 development due to lack of information on the primary endpoint(s) that are pursued during phase 3, due to limited data available then. For solid tumours, a commonly used early endpoint for efficacy is the objective response assessed using RECIST 1.1 criteria. These criteria are categorized into complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD) based on tumor measurements of target lesions, combined with categorical response of nontarget lesions and appearance of new lesions, where the measurements of target lesions are categorized into CR if target lesions disappear completely, PR if sum of diameters decreased by at least 30% from baseline (pre-treatment), PD if sum of diameters increased by at least 20% from the nadir (baseline or lowest before), or stable disease (SD) for all other scenarios. RECIST 1.1 have been used extensively in almost all solid tumor development as a gold standard for go-no-go decisions and have achieved great successes in oncology compound development. However, in the recent development of immunotherapies, people have found that some immunologic treatments may not produce high objective response rate (CR + PR based on RECIST 1.1 criteria) and still produce prolonged overall survival (OS). OS is a time-to-event endpoint that takes a long time to measure and needs many events to show the prolongation of OS. During the proof-of-concept trials, we usually won't be able to see many death events and we cannot wait until the OS is mature enough to assess the existence of OS signal. Due to these limitations, people have tried to assess if additional information can be obtained from tumor measurement data.

In this paper, we proposed new approaches to summarize tumor measurement data. In Section 2, we first discuss a new method to summarize tumor measurement data (to our knowledge, this method

was not used before). The method is called "**T**rajectory and non-missing **P**roportions **P**lot (TPP). The method focuses on the conditional mean response, condition on that the data are non-missing, and the proportion of non-missing, overtime. This plot can indeed be used for all trials to assess the treatment effect and non-missing patterns across all time points. These two parameters are jointly used to access the treatment effect and to answer questions about effectiveness. It will be seen that it requires both parameters to conclude that one treatment is better than other at a particular time point. The first parameter is obvious. The second parameter is from the belief that the proportion of non-missing from a more effective treatment (*A)* should not be less than that of a less effective treatment (*B)* to conclude that *A* is better than *B*. Another way to put this is: to conclude *A* is better than *B*, we should not know less about *A* than *B*.

In Section 3, we presented the method to combine the parameters from Section 2 and other efficacy parameters such as OS, PFS (progression-free survival), ORR (objective response rate), DOR (duration of response), TTR (time-to response) to make a well-informed go-no-go decision. In Section 4, we briefly discuss the application of the method presented in Section 2 in other therapeutic areas. Some including remarks are given in Section 5.

2. Trajectory and Proportion Plot

The number of patients enrolled in Phase 2 oncology trials are usually very limited. When OS is the primary endpoint of the subsequent studies, there aren't enough OS data to see a clear signal or convincing evidence. Additionally, we won't be waiting to see mature OS data if we see promising efficacy signals. Overall, we usually won't have reliable and matured OS data to make go-no-go decisions. This prompts people to look for other early endpoints for go-no-go decisions.

The data produced from RECIST 1.1 includes: ORR (objective response rate), TTR (time-to-response), DOR (duration of response), TTP (time-to-progression), and PFS (progression-free survival). Pembo data in the head and neck trial showed that even ORR of chemotherapy as the control is higher that pembro, Pembro beats the chemotherapy by OS. This example demonstrates that not all

efficacious treatments need to show ORR advantages over the control to win OS -- the clinical benefit endpoint. The question is: can tumor measurement data show additional signal that an efficacious treatment than the control so that an early go-no-go decision can be made?

- RECIST assessment data
    - Response: CR, PR, SD, PD
    - TTE (time to event) data such as: TTR, DOR, TTP.
    - TM data
- TB data are not commonly used directly, but are included in assignment of CR/PR/SD/PD only
- The belief is that "deep or durable tumor reduction" will produce prolonged PFS/OS.
    - Are 30% or 80% reduction in tumor size the same? Are a short-lived reduction and long-lasting one the same?
    - Due to high censoring rate in early phase PFS/OS data and the way to classify PD, some valuable information contained in TB data may be lost
        - Are 20% and 80% increase in tumor size the same? Are a rapid increase and slow one the same?

When TB data may have added value in predicting PFS/OS?

- Is PFS a "sufficient" statistic that summarizes all information in RECIST data?
- BOR, DOR and TTR have not summarized all information in RECIST data.
- Duration of SD and TTP for patients with BOR of SD and PD are not included in them.
- Rapid or slow increase in TB data are not captured by PFS
- This is perhaps one of the reasons that PFS sometimes does not predict OS.
- TB data have added predictive value: deep or durable reduction or slow increase in tumor burden may produce prolonged OS.
    - When censoring rate is high: predicting PFS, or OS
- Some treatments such as immunotherapy can be used after PD. In such scenarios, PFS2 may be used.

Pros and cons of some methods in summarizing TB data

- Mean TB using missing-at-random repeated measure analysis
  - TB data's missing mechanism is missing not at random. Hence, all such analyses are invalid.
- Analysis methods such as missing not-at-random methods require model assumptions. Any such assumptions may not be true.
- LOCF? Tumor keeps growing after patients drop out a trial. This analysis is fundamentally wrong.
- TB Slope
  - Pros: a single number. Easy to communicate.
  - Cons: It usually does not represent all information in TB data, slow but durable decrease won't be reflected as good one, compared with rapid but short-lived decrease and then rapid increase.

What does the Trajectory and Proportion Plot show?

- Denote:
  - *X(t):* the TB measurement of a patient at time *t*
  - *Y*: the total time a patient on study for TB measurement (time-to component)
  - Z(t): the indicator of a patient is censored for TB measurement or not. Z(t) = 1 indicating a patient is censored.
  - $\mu(t) = E(X(t)|Z(t)=0)$
- The plot gives estimates of $\mu(t)$ and Pr($Z(t)=0$) for all *t.*
- TB data has three components: X(t), Y, Z(t). It is perhaps impossible to summarize TB data by one statistic:
  - For oncology data, there must be a "time-to" component.
  - "deep" may be summarized by min(X(t))
  - "durable" information is buried in { *(X(t), Y, Z(t)),* all *t* }

Trajectory and Proportion Plot (TPP)

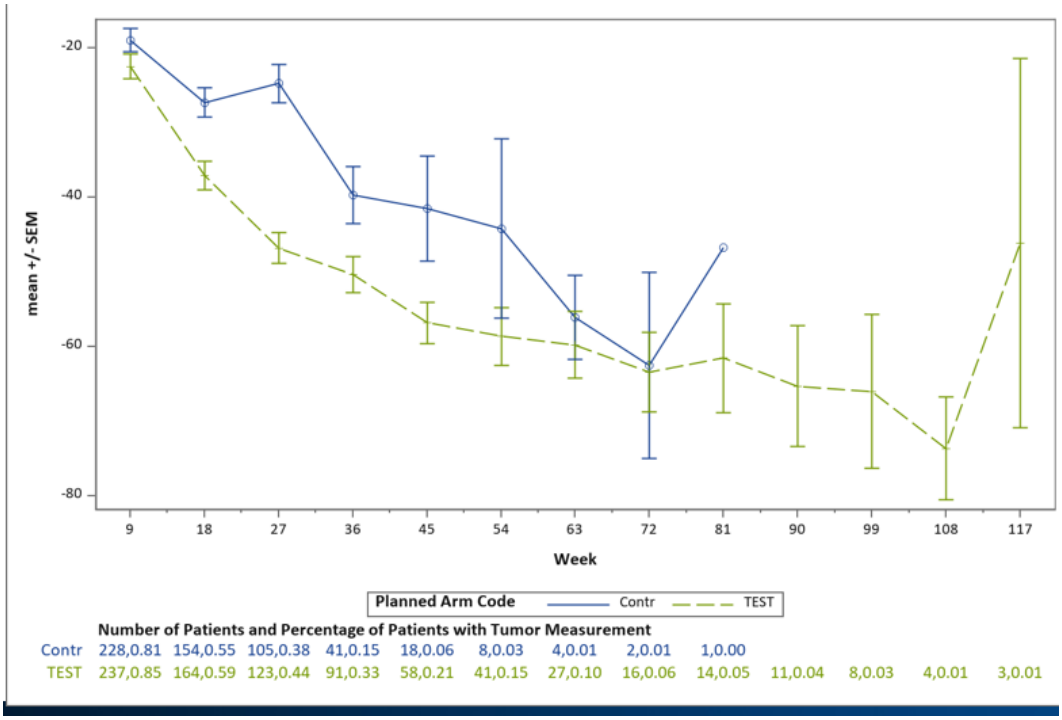
In the TPP below, it shows the mean and standard error of means over time, the percentage of the number of all patients with observed measurement at a particular timepoint over the number of patients at baseline (proportions), across all timepoints. Note that to conclude the test treatment has more tumor size reduction over the control at a particular time point *t*, there are two requirements:

1. The mean of the test treatment at *t* is less than that of the control;
2. The proportion of observed measurement at *t* is more than or equal to that of the control.

The first criterion is obvious. the second one can be concluded from the reasons of missing: in most cases, missing represents a no vote to the assigned treatment for lack of efficacy, intolerable toxicity, etc. Hence, the proportion of missing from the more effective treatment (*A)* should not be less than that of the less effective treatment (*B)* to conclude that *A* is better than *B*.

An example is given below to demonstrate that this TPP.

Trajectory and Proportion Plot



Note that though a common belief is that a deep and long tumor reduction will lead to clinical benefit, the addition of the criterion on proportion is needed to handle the role of patients without observation. Otherwise, the conclusion made only based on the trend of observed case is inherited biased and can be incorrect to all patients.

3. Totality of data recommended to use in decision-making process

A list of available data when a go-no-go decision is made is as follows:

- OS
- PFS
- OS2 = OS – PFS (to assess if there is carry-over effect)
- DOR
- ORR
- TTR

- ORR
- TB: percent change from baseline (or change from baseline)

How to assign scores by individual endpoint

- Assume that randomized phase 2 data are available. If not, assume historical data are available and can be used as the control for comparison.
- Choice of critical p-value for comparisons of OS, PFS, OS2, ORR: it depends on how many patients' data are available and importance of endpoints. Most likely only a trend can be seen but not at a high significance level. For example, $\alpha = 0.3$. The order of importance: OS > PFS > ORR > OS2.
- For DOR and TB, a trend can be assessed.
- For each endpoint, "PK" to assign a score to the test treatment. The test treatment scores
  - 1, if the test treatment wins over the control.
  - -1, if the test treatment loses to the control.

0, if the test treatment is similar to the control.

How to choose a winner?

- Simple summary of scores of the test and control treatments, respectively. Choose a winner based high score.
  - "win by number of endpoints". For example, 6, -6.
- Weighted scores based of the importance of these endpoints, relative to their registrational potential. For example,
  - OS: 0.3
  - PFS: 0.25
  - OS2: 0.1
  - DOR: 0.05
  - TTR: 0.05
  - ORR: 0.2
  - TB: 0.05
  - Interpretation
    - -1: loses by all endpoints (complete failure)
    - 1: wins by all endpoints
    - Between -1 and 1: wins by some and loses by some

Simulated data: Assessment

# of pts and cutoff after the last pts in	OS	PFS	OS2	DOR	TTR	ORR	TB	Score by # of wins	Score by weighted score
100, 6 months	1	1	0	1	0	1	0	4	0.8
100, 12 months	1	1	0	1	0	1	1	5	0.85
150, 6 months	1	1	0	1	0	1	0	4	0.8
150, 12 months	1	1	1	1	0	1	1	5	0.9
Full data	1	1	1	1	0	1	1	6	1

Note: If OS/PFS/OS2 do not show a trend due to high censoring rate and a compound does not improve ORR, DOR and TTR, then TB may be the only early indicator of improvement hence the only factor to make a go-decision.

4. General Use of TPP

The method of trajectory and proportion plot can be used generally to all other therapeutic areas.

5. Concluding Remarks

Recommendations

- Totality of data should be used when making go-no-go decisions, TB data should be part of the totality of data.
- TB data cannot be analyzed using repeated measurement analysis.
- TB data has two components, TB measurement and “time-to”. These two components reflect how “deep” and “durable” are and the speed of decrease/increase. These two measures need to be combined with proportions at each time point and across all time points for decision-making.
- The plot of mean and proportion of patients with observation across visits should be a common way to summarize trials with missing data: trajectory approach.
- The plot can be used for most of trials, not limiting to TB data.
- TPP is a general data summary method.