Chauhan Weighted Trajectory Analysis Reduces Sample Size Requirements and Expedites Time-to-Efficacy Signals in Advanced Cancer Clinical Trials

Utkarsh Chauhan 1, Daylen Mackey 2 and John R. Mackey 1,*

1 Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB T6G 2R7, Canada; uchauhan@ualberta.ca
2 CW Trial Analytics, Edmonton, AB T6E 4Y9, Canada; daylen.j.mackey@gmail.com
* Correspondence: jmackey@ualberta.ca

Abstract: (1) Background: As Kaplan–Meier (KM) analysis is limited to single unidirectional endpoints, most advanced cancer randomized clinical trials (RCTs) are powered for either progression-free survival (PFS) or overall survival (OS). This discards efficacy information carried by partial responses, complete responses, and stable disease that frequently precede progressive disease and death. Chauhan Weighted Trajectory Analysis (CWTA) is a generalization of KM that simultaneously assesses multiple rank-ordered endpoints. We hypothesized that CWTA could use this efficacy information to reduce sample size requirements and expedite efficacy signals in advanced cancer trials. (2) Methods: We performed 100-fold and 1000-fold simulations of solid tumor systemic therapy RCTs with health statuses rank-ordered from complete response (Stage 0) to death (Stage 4). At increments of the sample size and hazard ratio, we compared KM PFS and OS with CWTA for (i) sample size requirements to achieve a power of 0.8 and (ii) the time to first significant efficacy signal. (3) Results: CWTA consistently demonstrated greater power, and it reduced the sample size requirements by 18% to 35% compared to KM PFS and 14% to 20% compared to KM OS. CWTA also expedited time-to-efficacy signals by 2- to 6-fold. (4) Conclusions: CWTA, by incorporating all efficacy signals in the cancer treatment trajectory, provides a clinically relevant reduction in the required sample size and meaningfully expedites the efficacy signals of cancer treatments compared to KM PFS and KM OS. Using CWTA rather than KM as the primary trial outcome has the potential to meaningfully reduce the numbers of patients, trial duration, and costs to evaluate therapies in advanced cancer.

Keywords: Chauhan Weighted Trajectory Analysis; advanced cancer; Kaplan–Meier analysis; progression-free survival; overall survival

1. Introduction

Kaplan–Meier analysis with logrank testing is the gold-standard method for the analysis of efficacy outcomes in cancer trials. The KM estimator [1] has many advantages: it is a nonparametric method, making it suitable for analyzing time-to-event data that may not follow a normal distribution, and it can handle censored data effectively, where some individuals may not experience the event of interest by the end of the study, ensuring that these individuals are appropriately accounted for in the analysis. The KM estimator is typically used in conjunction with the logrank test [2], which allows for hypothesis testing to compare survival curves between different treatment groups in clinical trials. While the KM estimator provides a robust and widely accepted method for estimating survival functions from lifetime data, it does have several important limitations that reduce its utility. In particular, KM analyzes only a single time-dependent endpoint for each trial participant and cannot provide a single analysis that incorporates the many key efficacy endpoints that may occur during a cancer patient’s clinical course, including disease response, progression,
and death. Furthermore, KM analysis can only model unidirectional outcomes (for example, progression of disease) and cannot model bi-directional outcomes such as initial disease response, followed by later disease progression, and eventual death, that represent the typical illness trajectory of patients with advanced cancer in clinical trials.

The limitation of the current standard of KM-dependent trial designs is reflected in the extensive advanced cancer published literature, in which various individual clinical trial endpoints are assessed for their relative contributions to the actual clinical trajectory of the patient and their relative utilities as the single endpoint for studying new therapies [3–7]. This literature highlights a clear methodologic gap in existing trial design and analysis techniques.

Chauhan Weighted Trajectory Analysis (CWTA) [8] was specifically developed to address the limitations of KM analysis while retaining an intuitive visual and test statistic output that can be easily interpreted by clinical and regulatory communities due to its similarity to KM analysis. As a generalization of the KM methodology, CWTA has several advantages over traditional KM analysis. CWTA permits the assessment of outcomes defined by various ordinal grades or stages or clinical severity, which are common in medical settings but challenging to analyze using traditional methods like the KM estimator. Importantly, it facilitates continued analysis following changes in health state, which may be bidirectional (disease recovery or exacerbation), providing a more comprehensive view of the trajectory of clinical outcomes. Additionally, we proved the general case that using CWTA with multiple efficacy endpoints tended to reduce the samples sizes normally required for KM analysis to demonstrate a difference between treatment arms in a clinical trial [8].

CWTA also retains the merits of KM analysis. It is a non-parametric method with the ability to censor patients who withdraw or are lost to follow-up, ensuring robust analysis of clinical outcomes. CWTA, like KM, provides a graphical summary plot that visually depicts the trajectories of patients over time, making it easy to interpret and compare outcomes between different treatment arms. This is particularly important in the setting of clinical oncology, where oncologists have become accustomed to seeing a single plot for the efficacy endpoints of clinical trials, in which the separation of the control and experimental curves conveys meaningful information as to the differential effects of treatments through time. Finally, CWTA introduces a weighted logrank test, a modification of the traditional logrank test, to assess the statistical significance of differences in trajectories between groups, providing a rigorous method for hypothesis testing that is also a familiar analog for clinical oncologists.

Given these theoretical and practical advantages, we hypothesized that CWTA could productively be applied to the setting of RCTs in advanced cancer and could reduce sample size requirements and expedite efficacy signals when compared to standard KM analysis with a single efficacy outcome per patient.

2. Methods

2.1. Definitions

Progression-free survival (PFS) is an endpoint that is standardly defined in advanced cancer clinical trials as the time from random assignment in a clinical trial to disease progression or death from any cause. Overall survival (OS) is defined as the time from random assignment to death due to any cause. Partial response (PR), complete response (CR), stable disease (SD), and progressive disease (PD) were defined as per the RECIST 1.1 criteria [9].

2.2. Simulation Design

The simulation studies were generated using Python 3.8 [10]. The study simulations were stochastic processes in which randomly generated numbers were programmed to mirror the fluctuating solid tumour response to systemic therapy cycles with monthly measurements of treatment efficacy. We performed 100-fold and 1000-fold simulations of
RCTs in advanced cancer. RCTs were stochastically generated at a defined sample size (SS) allocated 1:1 to the control or experimental groups and run for 60 months. Rank-ordered health statuses were as follows: CR = 0, PR = 1, SD = 2, PD = 3, and Death = 4 (see Table 1). We assigned all patients to SD at time 0, and each month they were capable of a single-level transition, as specified in Table 1 as follows: either response (PR then CR), maintained SD, or irreversible exacerbation to PD or Death (see Figure 1 for example trajectories). Event probabilities were modified between groups as defined by a hazard ratio (HR). Through iterative programming of baseline transition probabilities, we modeled a control group CR rate of ~5% and a PR rate of ~30% to reflect first-line advanced cancer RCTs. We modeled a dropout rate of 10% that was uniformly distributed over 60 months.

Table 1. Ordinal variables used for advanced cancer clinical trial outcomes.

<table>
<thead>
<tr>
<th>Health Status</th>
<th>Ordinal Value</th>
<th>Type</th>
<th>Antecedent Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor in complete response (CR)</td>
<td>0</td>
<td>Reversible</td>
<td>Recovery after PR (1 → 0)</td>
</tr>
<tr>
<td>Tumor in partial response (PR)</td>
<td>1</td>
<td>Reversible</td>
<td>Recovery from SD (2 → 1) Exacerbation following CR (0 → 1)</td>
</tr>
<tr>
<td>Stable disease (at baseline; SD)</td>
<td>2</td>
<td>Reversible</td>
<td>Baseline condition Exacerbation following PR (1 → 2)</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>3</td>
<td>Non-reversible</td>
<td>Exacerbation of SD (2 → 3)</td>
</tr>
<tr>
<td>Death (all causes)</td>
<td>4</td>
<td>Non-reversible/</td>
<td>Exacerbation of PD (3 → 4)</td>
</tr>
</tbody>
</table>

Figure 1. Example disease trajectories. Patients were enrolled in simulated trials with stable disease (SD) and, through probabilities modified by hazard ratios and time since randomization, either experienced partial or complete response to systemic therapy (PR and CR), disease progression (PD), and death (D). Shown are the following three sample trajectories: in blue, a patient recovers to complete response but gradually worsens to progressive disease; in orange, a patient has stable disease aside from a brief partial response; in dark grey, a patient rapidly experiences disease progression and death.

2.3. Statistical Analysis

We performed Kaplan–Meier estimation using the Python 3.8 library “lifelines” [11]. This library was used to conduct logrank tests for KM PFS and OS. The results were validated by assessing the source code for accuracy and making a direct comparison to the results from SPSS v26 (IBM Corp., Armonk, NY, USA) [12]. CWTA was performed as described in Chauhan et al. 2023 [3] and augmented by a cloud-native cluster, with workloads executed in parallel across optimized Google Compute Engine virtual machines [13].
We ran 1000-fold analyses of completed trials programmed across sample sizes from 20 to 500 and hazard ratios from 0.5 to 0.8. For each trial, a p-value was computed using CWTA, KM PFS, and KM OS. The fraction of tests that were significant (at $\alpha < 0.05$) represents the power of the test (correctly rejecting the null hypothesis that the two groups are the same). At each hazard ratio, a sample size requirement for each of the three methods was interpolated as the threshold to reach 0.8 power, a common standard in RCT design.

Time to first significant efficacy was determined by calculating the mean and standard deviation of 100-fold and, for higher HR runs, 1000-fold simulations programmed across increments of sample sizes from 30 to 500 and hazard ratios from 0.5 to 0.8. Each simulated trial was incremented monthly up to 60 months and analyzed using CWTA (weighted logrank) and KM PFS and KM OS (logrank) to tabulate whether, at each monthly time point, statistical significance was reached ($\alpha < 0.05$). The trials that did not reach significance were omitted from the analysis. The analysis pipeline is summarized in Figure 2 (with contextual reference to Figures 1 and 3–5).

![Figure 2](image-url)  
Figure 2. A block diagram overview of the study methodology. Advanced cancer randomized controlled trials (RCTs) were stochastically generated at increments of defined sample size (SS) and hazard ratio (HR) allocated 1:1 to the control or experimental group. Health statuses were rank-ordered as per RECIST 1.1 criteria [9]. Each trial was analyzed using Kaplan–Meier progression-free survival (PFS) and overall survival (OS) using the logrank test and Chauhan Weighted Trajectory Analysis (CWTA) using the weighted logrank test. One-hundred- and one-thousand-fold simulations were performed to directly compare the three methods for power (by interpolation of sample size at 0.8 power) and time to study positivity (by incrementing time and assessing first time of $p < 0.05$). The diagram also outlines which figures in the manuscript correspond to different levels of analysis.

To assess how the output results varied with the input assumptions, sensitivity analyses were also performed with a control group CR rate of ~10% and PR rate of ~50% to reflect advanced cancer RCTs in which the control group had highly effective therapies.

3. Results

A single trial analyzed using KM OS, KM PFS, and CWTA is depicted in Figure 3. The simulations confirmed that CWTA was more powerful than KM analysis and reduced the required sample size to achieve 0.8 power across all the hazard ratios (Table 2, Figure 4). For
a typically sought hazard ratio of 0.8, CWTA reduced the sample size by 18% to 35% when compared to KM PFS across a range of hazard ratios and by 14% to 20% when compared to KM OS. The largest sample size reductions came in those trials with higher simulated hazard ratios (that is, less-effective interventions). We performed a sensitivity analysis with a higher efficacy control arm (Table S2, Figure S1), confirming similar sample size reductions (9% to 22%) and confirming that the largest reductions were seen in the higher simulated hazard ratios.

Figure 3. An identical advanced cancer trial analyzed three ways: using Kaplan–Meier (KM) overall survival, KM progression-free survival, and Chauhan Weighted Trajectory Analysis (CWTA). Input parameters included a sample size of 150, hazard ratio of 0.7, and a dropout rate of 10% uniformly distributed over a trial duration of 60 months. CWTA modeled partial and complete responses to therapy, and this added patient information facilitated greater divergence of plotted curves and statistical significance ($p = 0.026$).

Figure 4. Power as a function of sample size for Kaplan–Meier estimation of progression-free survival and overall survival and Chauhan Weighted Trajectory Analysis (CWTA). Using the assumptions of a moderately effective control arm, we performed 1000-fold simulations across a range of hazard ratios. CWTA demonstrated consistently higher power, reflecting a smaller sample size requirement during trial design. Assuming a trial designed to achieve a power of 0.8, CWTA reduced sample size requirements by 18% to 35% when compared to KM PFS and by 14% to 20% when compared to KM OS; the sample size reductions were most marked for studies designed for interventions with higher hazard ratios (that is, less-effective interventions).
Table 2. Eighty percent power sample size interpolation for Kaplan–Meier estimation of progression-free survival and overall survival and Chauhan Weighted Trajectory Analysis (CWTA) across a range of hazard ratios using moderately effective control arm assumptions. CWTA reduced the required sample size for 80% power by 18% to 35% versus KM PFS and by 14% to 20% versus KM OS.

<table>
<thead>
<tr>
<th>HR</th>
<th>CWTA</th>
<th>PFS</th>
<th>OS</th>
<th>Sample Size Reduction Using CWTA vs. PFS</th>
<th>vs. OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>54</td>
<td>66</td>
<td>63</td>
<td>18%</td>
<td>14%</td>
</tr>
<tr>
<td>0.6</td>
<td>69</td>
<td>100</td>
<td>83</td>
<td>31%</td>
<td>17%</td>
</tr>
<tr>
<td>0.7</td>
<td>130</td>
<td>192</td>
<td>162</td>
<td>32%</td>
<td>20%</td>
</tr>
<tr>
<td>0.8</td>
<td>317</td>
<td>486</td>
<td>395</td>
<td>35%</td>
<td>20%</td>
</tr>
</tbody>
</table>

When trials of a pre-specified sample size were run, CWTA markedly expedited the time-to-efficacy signals when compared to KM PFS and KM OS (Figure 5, Table S1A). This effect was seen across a broad range of sample sizes and hazard ratios and was confirmed in the sensitivity analysis with a highly active control arm (Figure S2, Table S1B). In the extreme cases with a highly effective therapy (low HR intervention), CWTA produced significant RCT outcomes in 20% of the time required to achieve a positive KM OS signal, although more typical was a 50% reduction in the time-to-efficacy signal for KM OS and a 40% reduction in the time-to-efficacy signal for KM PFS. The time to efficacy was most reduced in the studies with low hazard ratios (that is, highly effective interventions).

Figure 5. Time-to-efficacy signals comparison. We performed one-hundred-fold/one-thousand-fold simulations of time to study positivity as a function of sample size for Kaplan–Meier estimation of
progression-free survival and overall survival and for Chauhan Weighted Trajectory Analysis (CWTA) for a range of hazard ratios. CWTA demonstrated consistently shorter time to study positivity; this effect was most profound in studies with a highly effective intervention (that is, a low hazard ratio). (a) Time to study positivity at hazard ratio of 0.5; (b) time to study positivity at hazard ratio of 0.6; (c) time to study positivity at hazard ratio of 0.7; (d) time to study positivity at hazard ratio of 0.8.

4. Discussion

In cancer clinical trials, the evaluation of tumor response, progression-free survival (PFS), and overall survival (OS) are interrelated outcome metrics that collectively inform the assessment of the treatment efficacy. The solid tumor response, typically measured using the Response Evaluation Criteria in Solid Tumors (RECIST) [9], assesses the change in tumor size and the presence of new lesions as a direct response to systemic therapy. This metric is critical in the early phases (Phase I and IIb) of clinical trials to determine whether a drug has sufficient biological activity to warrant further study, and a strong response to therapy often correlates with a longer PFS [14]. PFS is valued in clinical studies, as it is considered a surrogate marker for therapeutic benefit, especially in settings where overall survival improvements might be difficult to demonstrate due to crossover designs and subsequent therapies. Studies have shown a correlation between prolonged PFS and increased overall survival [15]. While these relationships generally hold true, the specific relationships amongst response, PFS, and OS may depend on the tumor or tumor subtype studied, the line of therapy, and the availability of effective salvage regimens on progression [16,17].

In this study, we performed simulations of therapeutic trials conducted in the setting of advanced cancer and analyzed the studies with the traditional KM methodology as well as with its novel generalization, CWTA. By virtue of the ability of CWTA to incorporate and analyze multiple rank-ordered health status endpoints that fluctuate, in this specific case, including complete response, partial response, stable disease, disease progression, and death, CWTA outperformed the traditional KM analysis for both PFS and OS. These findings were robust across a representative range of trial sample sizes and across the range of effect sizes typically sought after and observed in advanced cancer trials.

The most dramatic demonstration of the improved performance of CWTA over KM was the meaningful reduction in the sample size requirements to achieve the typical power threshold of 0.8; sample size reductions ranging from 9% to 35% were noted across the spectrum of the study designs (Figure 4, Table 2, Figure S1, Table S2). In the setting of advanced cancer, smaller sample requirements are very desirable. Smaller trials are less costly and accrue and complete more rapidly. Increasingly, molecularly targeted subpopulations of cancer patients are being studied, where the drug target may have a prevalence of only a small percentage of the population, and large trials become unfeasible. Similarly, some cancer populations have multiple competing open studies, and sample sizes become the critical determinants of a study’s chance of success. Proving or disproving efficacy in a smaller population of study participants also spares ineffective therapy and unnecessary follow-up.

There is a direct relationship between the sample size and cost of an advanced cancer clinical trial. As the sample size increases, the cost of the clinical trial also increases due to direct increases in the costs for each of the following: (i) participant recruitment outreach efforts, screening processes, and incentives; (ii) monitoring and data collection resources for monitoring participant health and ensuring adherence to the trial protocol; (iii) logistics and infrastructure for handling the participants, including additional clinical sites and staff; and (iv) treatment, on-treatment care, and long-term follow-up. A study published by the Tufts Center for the Study of Drug Development demonstrated that each additional patient in a phase III oncology trial can increase the cost by approximately USD 36,000 [18].

Another demonstration of the improved performance of CWTA over KM manifested as the statistically and clinically significant improvements in the time-to-efﬁcacy signals generated by the study, analyzed with both methodologies. The magnitude of these
benefits (2- to 6-fold faster time to study positivity; see Figure 5 and Table S1) has pragmatic implications for the conduct of trials in advanced cancer. These data suggest that interim and primary analyses of cancer trials can be conducted earlier in the course of the study, leading to a more rapid determination of futility or study benefit. In this way, ineffective therapies can be identified, permitting cross-over to better options, and effective therapies can be found more expeditiously. Early recognition translates directly into improved ethical standards and quality of life of enrolled subjects. In aggregate, more timely reporting of new, effective interventions accelerates the time to population benefit. Furthermore, because many of the inputs into clinical trial costing represent time-dependent and relatively fixed costs throughout the study, reductions in time-to-efficacy reporting in a clinical trial provide additional dramatic reductions in study costs [19].

The rapid expansion in the number of new therapeutic agents in oncology has created an urgent need for more efficient clinical trial processes. Traditional trial designs based on KM analysis are often slow and resource-heavy, making them inadequate for keeping pace with the rapid advancements in drug development. Adaptive trial designs, such as basket trials and umbrella trials, provide a partial solution by enabling the evaluation of multiple therapies or therapy combinations simultaneously, thus potentially reducing the time and cohort size needed to determine a drug’s efficacy. Basket trials test the efficacy of one drug across different cancer types that share a common mutation [20], while umbrella trials test multiple drugs in a single cancer type but against different mutations [21]. In principle, CWTA could readily be incorporated into such studies to further reduce sample sizes and expedite an efficacy signal.

In any modeling exercise, some simplifications are required. For this study, all the patients were accrued simultaneously, and the fully accrued trial was run from time 0, without simulating the staggered accrual that typically occurs in advanced cancer trials. However, sensitivity testing suggests that time to accrual does not meaningfully change the magnitude of these findings. In our simulations, we tested each month for statistical significance, and we did not adjust for multiple testing; while there was a systematic bias toward underestimating the time-to-positive study results, this bias was shared across each of the three analyses we conducted (CTWA, KM PFS, and KM OS), with the relative time-to-positive signals being the key output rather than the absolute times. In practice, such interim and final analyses would typically be pre-specified in the case of KM PFS and KM OS by pre-specified events or, in the case of CWTA, specified by a pre-defined number of health state transitions.

These data provide the conceptual framework for interrogating completed studies, in which datasets from advanced cancer trials could be studied to confirm the results of time to study positivity using actual trial data sequentially analyzed at similar intervals. Furthermore, in principle, the improvements in both the sample size requirements and time-to-study-efficacy signal should be evident in the setting of early-stage cancer trials; we are pursuing additional simulation work to explore this possibility.

5. Conclusions

In aggregate, this study shows that by incorporating the entirety of the cancer trajectory, including disease response, progression, and death, CWTA meaningfully reduces the sample size requirements for clinical trials in the advanced cancer setting, particularly in studies of less-effective interventions. Furthermore, CWTA expedites the time-to-efficacy signals of cancer treatments compared to KM PFS and OS. Consequently, these data suggest that using CWTA rather than KM as the primary trial outcome has multiple efficiency advantages and the potential to reduce the trial duration, cost, and numbers of patients needed to evaluate therapies in advanced cancer.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biomedinformatics4030092/s1, Figure S1: Power as a function of sample size for Kaplan–Meier estimation of progression-free survival and overall survival and Chauhan Weighted Trajectory Analysis (CWTA) using a highly active control regimen; Figure S2:
Time-to-efficacy signals comparison using a highly active control regimen; Table S1: Time-to-efficacy signals comparison for (A) moderately active control regimen and (B) highly active control regimen; Table S2: 80% Power sample size interpolation for Kaplan–Meier estimation of progression-free survival and overall survival and Chauhan Weighted Trajectory Analysis (CWTA) across a range of hazard ratios using a highly active control regimen.

**Author Contributions:** Conceptualization, U.C. and J.R.M.; methodology, U.C. and J.R.M.; software, U.C. and D.M.; validation, U.C. and J.R.M.; formal analysis, U.C. and J.R.M.; investigation, U.C.; resources, U.C. and D.M.; data curation, U.C. and D.M.; writing—original draft preparation, U.C. and J.R.M.; writing—review and editing, D.M.; visualization, U.C. and J.R.M.; supervision, J.R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of the research are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Conflicts of Interest:** J.R.M. reports leadership roles and equity in SMHeartCard Inc, Pacylex Pharmaceuticals, and IllumiSonics, Inc. D.M. and U.C. hold equity in CW Trial Analytics.

**References**


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.