The Importance of the Change Point Recognition in the Landmark Analysis of Immuno-oncology Clinical Trial with the Delayed Treatment Effect

Lihong Huang (✉ huang.lihong@zs-hospital.sh.cn)

Zhongshan Hospital Fudan University

Naitee Ting

Boehringer Ingelheim Pharmaceuticals Inc

Research article

Keywords: Immuno-oncology, Clinical trials, Delayed clinical effect, Time-to-event outcome

Posted Date: August 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-48377/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The proportional hazards (PH) assumption is often not met in the IO clinical trials due to the delayed effect. The landmark analysis could be performed as sensitivity analysis to evaluate the efficacy of the treatment despite the impact of the violation of PH assumption, but analysis results can vary according to the different landmarks. The goal of the paper was to raise the awareness in both the statistical and clinical communities regarding the importance of the change point in the IO delayed effect clinical trials, in order for the improvement of the existing landmark analysis method.

Methods: Pre-define the change point as an objective choice of landmark time of landmark analysis was recommended for the IO trials with delayed effect to avoid additional biases caused by the choice of landmark. We called this method change point landmark analysis (CPLA). Monte Carlo simulation was implemented to explore test power and type I error of CPLA. A simulated example was also conducted to evaluate the advantages of change point recognition on the landmark analysis.

Results: CPLA showed the high power comparing to the earlier landmark choices and the later landmark choices, especially when the impact of the delayed effect was earlier than the median survival time. The type I errors of CPLA are also well controlled.

Conclusions: the change point is a good objective choice of landmark in landmark analysis, which could resolve the objectively pre-defined challenge which is the critical requirement of the landmark analysis.

Background

In recent years, due to the innovative medical advancement, new treatments such as immunotherapies have yielded long term survivors. The development of immuno-oncology (IO) new drugs or the investigation of new indications for the IO anti-cancer drugs already on the market have become the drug research hotspots.

A time to event endpoint is the primary outcome in many clinical trials, especially oncology drug development. The most commonly used statistical methods for time to event analyses have been the log-rank test[1] and Cox regression[2]. In both cases, the performance of analysis depends on the proportional hazards (PH) assumption. Generally speaking, the PH assumption is often not met in IO clinical trials. IO agents can have effect on both the human immune system and the tumor microenvironment, the effect of IO agent is not typically directed to the tumor itself, it instead boosts or releases the brake from the patient immune system, and this positive effect may not be observed immediately, therefore the delayed separation with a change point (CP) of the Kaplan-Meier curves[3] have often been observed in the IO therapies trials, potentially resulting in a violation of PH assumption[4–6]. The published simulation study showed that the conventional study design with exponential assumption could lead to an underestimation of statistical power in the presence of delayed clinical effect when designing randomized clinical studies with immunotherapies[4].

Therefore, the delayed effect type survival curve observed in the IO clinical trials, consequently pose unique challenges to the trial design and analysis method. Can we recognize the change point when survival curves cross in IO clinical trials and specify the optimal development strategy? How can we make use of this kind of delayed effect with change point information for the statistical analysis?

In the survival analysis setting, landmark analysis refers to the practice of designating a time point occurring during the follow-up period (known as the landmark time) and analyzing only those subjects who have survived until the landmark time[7]. That is to say, Once the landmark has been chosen, any ineligible subjects would have been excluded, and subjects have been classified according to their status at the landmark time, the usual survival analysis methods are applied, and the conclusions are only generalizable to subjects who have survived until the landmark time. Analysis results can vary according to the different landmarks. The hazard ratio estimations are conditional with a different target at each landmark. Therefore, the choice of landmark is a critically important consideration. To avoid additional biases caused by the choice of landmark, the landmark should be selected a priori, based on some clinically significant natural time before the start of data analysis[8].

The goal of the paper was to raise the awareness in both the statistical and clinical communities regarding the importance of the change point in the IO delayed effect clinical trials, in order for the improvement of the existing landmark analysis method.

Methods

The Change Point of the Survival Curves

A change point is the location where the distribution abruptly changes in a data sequence [9]. Due to the mechanism of the IO anti-cancer drugs, the delayed separation of Kaplan-Meier survival curves is usually observed, sometimes the curves are converged before the change point (Figure1A), sometimes the curves are very close (Figure1B) before the change point. When the delayed separation of survival curve is present, it violates the fundamental study design assumption of the proportional hazards, and also results in a potential loss of statistical power to demonstrate the difference between two treatment groups because of the long invalid period before the separation change point.

Landmark analysis

With the exponential distribution assumption, the survival function is defined as:

$$\lambda(t) = \lambda_0 \exp(\beta \text{group} \times t_g)$$

$$t_g = I(t \geq t_{CP}) = \begin{cases} 0 & t < t_{CP} \\ 1 & t \geq t_{CP} \end{cases}$$

where λ_0 , $group$ and t_g are baseline hazard, treatment group and the time point indicator, t_{CP} is the time of change point. The hazard ratio before change point and after change point under the exponential distribution in Figure 1A should be estimated separately as:

$$\begin{aligned}
 group = 1: \quad \lambda_1(t) &= \lambda_0 \exp(\beta t_g) \\
 group = 0: \quad \lambda_0(t) &= \lambda_0 \\
 HR = \frac{\lambda_1(t)}{\lambda_0(t)} &= \exp(\beta \times t_g) = \begin{cases} 1 & t_g \leq t_{CP} \\ \exp(\beta) & t_g > t_{CP} \end{cases}
 \end{aligned}$$

For the landmark analysis, only the subjects survived in the period $t_g > t_{CP}$ would be involved in the analysis. We recommend to pre-define the change point as an objective choice of landmark time of landmark analysis for the IO trials with delayed effect to avoid additional biases caused by the choice of landmark. We called this method change point landmark analysis (CPLA) in our research, and then change point was named as change point landmark in CPLA.

Simulation studies

Two simulation studies were conducted in our research.

Power and type I error evaluations

The first simulation study was performed to evaluate test power (scenarios A and B) and type I error (scenario C) of the change point landmark analysis. We conducted Monte Carlo simulations for landmark analyses and traditional full data Cox regression using delayed effect type survival data with change point at $CP=2.5$ month, three landmarks 1, 2.5 and 3.5 months were considered, one was before the true change point, one was after the true change point. Equal sample sizes ($N_1=N_2=20,30,40$) simulated data were generated. A simulation of time-to-event data was performed based on a randomly censored model [10, 11], we generated individual lifetime X following the survival functions in Table1. To test the power of landmark analyses, the scenarios A were the simulated situations for the survival curves as Figure 1A, and the scenarios B were for Figure 1B. Different median survival times, different hazard ratios in the before and after change point periods were also considered in the power tests. Additionally, the scenario C was two overlapping survival curves to test type I errors. The censoring time S in two samples was generated from uniform distributions $U(0, a)$ and $U(0, b)$, where varying the values of a and b may result in censoring rates of approximately 10%, 20%, or 30% in the two samples. Because the lifetime X followed different distributions in each group, it was necessary for the values of a and b to be unequal to keep the average censoring rates in each group approximately equal to the given censoring rates. Each individual was assigned an observed survival time $T=\min(X, S)$ and an event indicator $\Delta=I[X \leq S]$. The exact power and size of test statistics were estimated by determining the proportion of samples for which the null hypothesis was rejected at the $\alpha = 0.05$ significance level, based on 1000 simulations.

We performed simulations of trial designs to evaluate the advantages of change point recognition on the landmark analysis. Our initial simulations use the similar scenario of Chen[4] and Korn[12]: 680 patients are accrued and randomly assigned 1:1 over 34 months and final analysis occurred at 48 months after the first patient was randomly assigned. If there is no delay in treatment effect, this design has 90% power to detect a hazard ratio (HR) of 0.75 (using a two-sided 0.05-level log-rank test) with the 512 events in the final analysis, and there is no loss to follow-up. As is well known, the sample size and study duration are always fixed according to the sponsor's budget during the study design, then we fixed the total duration as 48 months accordingly. Because it is also hard to know the actual treatment effect delay length in advance, to assess the impact of the delayed clinical effect, we considered the 3 scenarios which covered 1/12, 1/8, 1/6 delay of the total study duration, then the change points were at the 4th month, the 6th months and the 8th months (Figure 2), where the hazard ratio was 1.0 before the change points and 0.75 thereafter. As the whole study duration was pre-defined as 48 months, the observed events were decreased for the impact of the delayed effect. No interim analysis was considered in the simulated trials. Each scenario was evaluated based on 1000 simulations, the power and mean of the conditional HRs with 95% confidence interval (CI) were calculated for change point landmark, other landmarks and full data Cox regression.

Results

In the first simulation study, comparing the power of landmark analyses when predefined landmarks as 2.5 months, 1 month and 3.5 months, the CPLAs (2.5 months) in the scenarios A have the highest power despite the impact of sample sizes and censor rates (Table 2), even when the change point is close to the median survival time 2.8 months in the scenarios A1, or when the HR is 0.8 after change point in the scenarios A3. The P values distribution of the change point landmark (2.5 months) is also distinct lower than the other landmarks from 0 to 6 months (Figure 3). In scenario B1, when the change point is close to the median survival time, the power of the CPLAs does not show the stable highest power. But In scenarios B2 and B3, when the change points have a certain distance from the median survival time, the CPLAs keep the highest power comparing to the other landmarks, even when the HR is little bit larger in scenario B3 (Table 3).

Meanwhile, the power of Cox regression is obvious lower as the PH assumption is violated by the delayed effect (Table 2 and Table3). On the other hand, the type I errors of the change point landmark analyses are well controlled, as well as the other landmarks (Table 4).

In the second simulated clinical trial, although the sample size and the event numbers in the final analyses are lower than the designed setting due to the delayed effect under fixed study duration, the power of true change point landmarks are still around 90%, but the estimations of conditional HR are slightly higher than 0.75 – the pre-specified value. The earlier landmarks (e.g. 4 months landmark in 6 months delayed setting, 4 months and 6 months landmarks in 8 months delayed setting) have the

obvious higher power than the later landmarks, but still lower than the true change point landmarks (Table 5).

Discussion

The possibility of a delayed treatment effect has become more relevant to IO cancer trials, and several trials have been demonstrated to show that[13, 14], like the checkpoint inhibitors (anti-cytotoxic T-lymphocyte antigen 4 [CTLA-4] and inhibitors to programmed death 1 [PD-1] or its ligand [PD-L1]. Biologic rationale and empirical evidence suggest that these immune agents may take a few months to show benefit[15, 16]. The potential loss of power, especially the effects of futility interim monitoring also has been noted[17, 18] in IO cancer trials. Recently, landmark analyses were performed to explore the association of treatment and long-term clinical outcomes[19], and were commonly used in the melanoma clinical trials[20, 21]. As reported, landmark analysis is mentioned and the original article cited in more than 420 papers, including the journal *Circulation*, *The lancet*, *The New England Journal of Medicine* and *Journal of Clinical Oncology*[8].

In this article, the landmark analysis was recommended to be applied in IO cancer trials with delayed effects and crossed survival curves. But the choice of landmark time is very critical because this method omits the time-to-event distribution before landmark time, different landmark times estimate different conditional hazard ratios and may suffer from loss of power with misclassification[8]. Therefore, the arbitrary choice of landmark may be questioned[22]. We recommended to use the change point of the survival curves for the objective choice of the landmark for the IO cancer trials.

From simulation studies, CPLA showed the highest power comparing to the earlier landmark choices and the later landmark choices, especially when the impact of the delayed effect was earlier than the median survival time. Interestingly, it was reported that the landmark method is most powerful when the “risk altering intervening event” occurs comparatively early and the outcomes of interest are not particularly common at this early study point[8], the recognition of the change point of the delayed effect in the IO cancer trials is a useful approach to figure out the earliest timepoint of the “risk altering intervening event” from statistical point of view. Furthermore, the landmark should be selected in advance to safeguard against the danger of a data-driven decision, the change point could be objectively pre-defined as the landmark in the statistical analysis plan. Therefore, it is worth to figure out the change point objectively based on the crossed delayed effect survival data.

Statistical methodology such as weighted log-rank test (WLR), Weighted Kaplan-Meier test et al [23, 24], also exist to account for the delayed separation of Kaplan-Meier curves. However, landmark analysis is easy to execute and understandable for the clinicians, this method keeps having its superiorities and applied values.

It is worth noting that the CPLA may not show the treatment advantages if the change point is near to the median survival time, but this kind of longer delayed effect is not common in the IO clinical trials. If there is evidence supporting a very long delayed effect, it is better to consider that as the part of the

design of the subsequent trial, for example, randomizing the treatment assignment after the delayed effect period of the treatment.

Conclusions

The change point is a good objective choice of landmark (1) the CPLA have the highest power when the change point is not near to the mediation survival time, (2) the change point can resolve the objectively pre-defined challenge which is the critical requirement of the landmark analysis.

Abbreviations

PH

proportional hazards; IO:immuno-oncology; CP:change point; CPLA:change point landmark analysis; HR:hazard ratio; CI:confidence interval;

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not Applicable

Availability of data and materials:

The simulated data is available on request.

Competing interest:

The authors have no conflicts of interest to declare.

Funding:

This work was supported by the National Natural Science Foundation of China (Grant 81903407 to Dr. Huang). The funding source had no role in the design of the study and its execution, analyses, interpretation of the data, or writing the manuscript.

Author's contributions:

LH conducted literature review and final data analysis, created tables and figures, wrote methods, results, discussion and conclusion sections. NT revised and edited all drafts of the manuscript. All authors read and approved the final manuscript.

Acknowledgments:

Not Applicable

References

1. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*. 1972;135:185–98.
2. Cox DR. Regression models and life-tables. *J Roy Stat Soc B*. 1972;34:187–202.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*. 1958;53:457–81.
4. Chen T-T. Statistical issues and challenges in immuno-oncology. *J immunother Cancer*. 2013;1:18.
5. Kaufman PA, Awada A, Twelves C, et al. Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *Journal of clinical oncology*. 2015;33:594.
6. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373:1627–39.
7. Wills MA. Morphological disparity: a primer. *Fossils, phylogeny, and form*: Springer, 2001:55–144.
8. Dafni U. Landmark analysis at the 25-year landmark point. *Circulation: Cardiovascular Quality Outcomes*. 2011;4:363–71.
9. Chen H, Zhang N. Graph-based change-point detection. *The Annals of Statistics*. 2015;43:139–76.
10. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Statistics in medicine*. 2013;32:4118–34.
11. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statistics in medicine*. 2006;25:4279–92.
12. Korn EL, Freidlin B. Interim futility monitoring assessing immune therapies with a potentially delayed treatment effect. *J Clin Oncol*. 2018;36:2444.
13. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363:711–23.
14. Wolchok JD, Neyns B, Linette G, et al. Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. *The lancet oncology*. 2010;11:155–64.

15. Finke LH, Wentworth K, Blumenstein B, Rudolph NS, Levitsky H, Hoos A. Lessons from randomized phase III studies with active cancer immunotherapies–outcomes from the 2006 meeting of the Cancer Vaccine Consortium (CVC). *Vaccine*. 2007;25:B97–109.
16. Anagnostou V, Yarchoan M, Hansen AR, et al. Immuno-oncology trial endpoints: capturing clinically meaningful activity: AACR, 2017.
17. Menis J, Litière S, Tryfonidis K, Golfopoulos V. The European Organization for Research and Treatment of Cancer perspective on designing clinical trials with immune therapeutics. *Annals of translational medicine* 2016; 4.
18. Xu Z, Zhen B, Park Y, Zhu B. Designing therapeutic cancer vaccine trials with delayed treatment effect. *Statistics in medicine*. 2017;36:592–605.
19. Eisenstein EL, Anstrom KJ, Kong DF, et al. Clopidogrel use and long-term clinical outcomes after drug-eluting stent implantation. *Jama*. 2007;297:159–68.
20. Weber JS, Hodi FS, Wolchok JD, et al. Safety Profile of Nivolumab Monotherapy: A Pooled Analysis of Patients With Advanced Melanoma. *J Clin Oncol*. 2017;35:785–92.
21. Ascierto PA, Long GV. Progression-free survival landmark analysis: a critical endpoint in melanoma clinical trials. *The Lancet Oncology*. 2016;17:1037–9.
22. Liang F, Zhang S, Zhu J. Problematic Landmark Analysis Has Led to a Problematic Conclusion. *J Clin Oncol*. 2017;35:1967–8.
23. Li H, Han D, Hou Y, Chen H, Chen Z. Statistical inference methods for two crossing survival curves: a comparison of methods. *PLoS One*. 2015;10:e0116774.
24. Xu T, Zhu D. A review of statistical methods on testing time-to-event data. *Biometrics Biostatistics International Journal*. 2018;7:570–2.

Tables

Table 1. Simulation parameters for the time-to-event data

Scenarios	Survival Curves	<i>Median Survival*</i> [Before CP]	<i>Median Survival*</i> [After CP]	<i>HR</i> (Before CP)	<i>HR</i> (After CP)
A1	1	2.8m	2.8m	1.00	0.67
	2	2.8m	4.2m		
A2	1	5.6m	5.6m	1.00	0.67
	2	5.6m	8.2m		
A3	1	5.6m	5.6m	1.00	0.80
	2	5.6m	6.9m		
B1	1	2.8m	2.8m	1.33	0.67
	2	2.1m	4.2m		
B2	1	5.6m	5.6m	1.33	0.67
	2	4.2m	8.2m		
B3	1	5.6m	5.6m	1.33	0.80
	2	4.2m	6.9m		
C	1	2.8m	2.8m	1.00	1.00
	2	2.8m	2.8m		

m: month; CP: Change point, which is 2.5m; * relative to time 0.

Table 2. Power of Cox regressions with different landmarks (scenario A)

Scenarios	Censor	Sample size	Landmark T=1m		Landmark T=2.5m		Landmark T=3.5m		Full Data	
			n_1	Rejection (%)	n_2	Rejection (%)	n_3	Rejection (%)	N	Rejection (%)
A1 (Power)	10%	$N_1=N_2=20$	30	28.3	20	67.3	17	33.9	40	12.3
	20%		30	27.8	18	58.9	14	26.6	40	11.3
	30%		29	22.6	15	27.8	11	7.7	40	9.0
	10%	$N_1=N_2=30$	46	57.6	29	91.9	25	55.9	60	30.1
	20%		45	55.1	27	88.7	21	53.2	60	27.1
	30%		43	47.3	23	55.4	16	28.0	60	19.6
	10%	$N_1=N_2=40$	61	79.4	40	98.4	33	68.8	80	50.8
	20%		60	76.4	36	96.7	28	69.4	80	47.5
	30%		58	65.4	31	72.3	21	47.8	80	35.3
A2 (Power)	10%	$N_1=N_2=20$	34	36.6	28	61.9	25	37.1	40	24.7
	20%		34	38.4	26	62.8	22	38.5	40	24.5
	30%		33	31.2	23	51.6	19	28.6	40	18.4
	10%	$N_1=N_2=30$	52	69.2	41	91.4	38	62.9	60	54.7
	20%		51	66.8	39	89.9	34	62.5	60	48.9
	30%		50	59.3	35	84.6	29	53.8	60	39.9
	10%	$N_1=N_2=40$	70	90.4	55	98.8	50	78.0	80	78.4
	20%		68	86.9	51	97.1	45	75.4	80	69.4
	30%		66	78.7	47	95.2	38	69.9	80	59.1
A3 (Power)	10%	$N_1=N_2=20$	34	9.0	28	18.7	24	14.2	40	4.3
	20%		34	11.8	26	27.5	22	18.5	40	6.2
	30%		33	11.3	23	28.0	19	17.1	40	5.7
	10%	$N_1=N_2=30$	52	23.1	41	48.4	37	29.4	60	12.0
	20%		51	29.7	39	58.8	33	37.4	60	16.0
	30%		50	28.6	35	57.9	28	38.4	60	14.1
	10%	$N_1=N_2=40$	70	47.3	55	75.7	49	45.6	80	28.4
	20%		68	52.3	51	79.6	44	55.2	80	33.3
	30%		66	48.5	47	78.3	37	51.3	80	27.5

Table 3. Power of Cox regressions with different landmarks (scenario B)

Scenarios	Censor	Sample size	Landmark T=1m		Landmark T=2.5m		Landmark T=3.5m		Full Data	
			n ₁	Rejection (%)	n ₂	Rejection (%)	n ₃	Rejection (%)	N	Rejection (%)
B1 (Power)	10%	N ₁ =N ₂ =20	29	25.9	18	66.6	14	49.0	40	3.6
	20%		29	16.6	16	7.3	12	5.1	40	2.5
	30%		28	5.4	13	0.0	9	0.0	40	0.4
	10%	N ₁ =N ₂ =30	44	44.9	26	93.1	22	82.5	60	7.8
	20%		43	32.7	24	19.0	18	15.6	60	4.9
	30%		41	13.1	20	0.0	13	0.0	60	0.2
	10%	N ₁ =N ₂ =40	58	62.9	35	98.2	29	94.6	80	15.2
	20%		57	50.2	31	31.0	23	29.0	80	7.4
	30%		55	23.5	27	0.0	17	0.0	80	0.8
B2 (Power)	10%	N ₁ =N ₂ =20	34	40.3	26	82.7	24	62.2	40	17.7
	20%		33	37.9	24	72.0	21	55.8	40	16.5
	30%		32	28.0	22	44.0	18	34.0	40	8.0
	10%	N ₁ =N ₂ =30	51	69.8	39	98.2	36	90.2	60	38.3
	20%		50	64.2	36	94.2	31	84.7	60	30.8
	30%		48	53.0	33	68.9	27	58.6	60	19.3
	10%	N ₁ =N ₂ =40	68	88.2	52	99.8	47	97.4	80	59.2
	20%		66	82.0	48	98.8	42	96.0	80	51.0
	30%		65	69.9	44	82.1	36	77.8	80	32.2
B3 (Power)	10%	N ₁ =N ₂ =20	34	11.9	26	48.3	24	23.7	40	2.8
	20%		33	16.3	24	56.9	21	31.5	40	4.1
	30%		32	14.9	22	47.7	18	25.5	40	3.6
	10%	N ₁ =N ₂ =30	51	31.1	39	82.5	36	51.9	60	7.6
	20%		50	36.3	36	88.0	31	61.6	60	9.5
	30%		48	31.8	33	83.8	27	57.1	60	7.7
	10%	N ₁ =N ₂ =40	68	53.8	52	94.9	47	74.5	80	16.2
	20%		66	56.8	48	97.1	42	79.1	80	19.8
	30%		65	49.6	44	95.3	36	77.7	80	14.8

Table 4. Type I error of Cox regressions with different landmarks (scenario C)

Scenario	Censor	Sample size	Landmark T=1m		Landmark T=2.5m		Landmark T=3.5m		Full Data	
			n ₂	Rejection (%)	n ₁	Rejection (%)	n ₃	Rejection (%)	N	Rejection (%)
C (Type I Error)	10%	N ₁ =N ₂ =20	26	5.3	19	4.9	16	4.6	40	4.1
	20%		26	6.0	19	5.9	16	4.0	40	4.8
	30%		26	6.6	19	4.0	16	4.8	40	6.2
	10%	N ₁ =N ₂ =30	39	4.4	28	5.5	24	5.8	60	4.9
	20%		39	6.0	28	6.0	24	5.6	60	5.3
	30%		39	5.3	28	5.2	24	5.2	60	5.6
	10%	N ₁ =N ₂ =40	52	4.2	38	5.2	32	4.4	80	5.1
	20%		52	5.5	38	5.2	32	5.1	80	5.5
	30%		52	4.0	38	5.3	32	5.2	80	4.4

Table 5. Change point estimation and landmark analyses for the simulated trial

Delayed effect	Models	n	Events	Power (%)	Conditional HR (95% CI)
4 months	Change point Landmark	590	413	89.1	0.80(0.78, 0.83)
	Other landmarks:				
	6 months	565	387	5.6	0.86 (0.81, 0.91)
	8 months	531	353	1.6	0.87 (0.82,0.93)
	Cox regression	680	502	60.2	0.84 (0.81, 0.86)
6 months	Change point Landmark	550	372	91.9	0.79 (0.76, 0.82)
	Other landmarks:				
	4months	590	413	83.7	0.81 (0.78, 0.84)
	8 months	531	353	1.6	0.87 (0.82, 0.93)
	Cox regression	680	502	48.6	0.84 (0.81, 0.87)
8 months	Change point Landmark	512	334	93.4	0.78(0.74, 0.81)
	Other landmarks:				
	4 months	590	413	72.5	0.81 (0.78, 0.85)
	6 months	550	372	85.3	0.80 (0.77,0.83)
	Cox regression	680	502	34.6	0.85 (0.82, 0.87)

Figures

Figure 1

Graphical presentation of Kaplan-Meier survival curves with delayed clinical effect. Delayed clinical effect with the completed coincident survival curves before the change point; Delayed clinical effect with the close survival curves before the change point.

Figure 2

Simulated survival curves with different delayed effects

Figure 3

Landmark analyses P values for scenario A2 (simple size $N_1=N_2=40$, censor rate 10%)